Integrating Network Digital Twinning into Future AI-based 6G Systems

D2.1

Data governance, privacy, and harmonization

Document Information	
Grant Agreement N°	101136314
Authors and institutions	Ayat Zaki Hindi (LIST), Jean-Sébastien Sottet (LIST), Ion Turcanu (LIST), Sébastien Faye (LIST), Ayse Sayin (EBY), Ipek Arikan (EBY), Ramin Fuladi (EBY), Ana Pereira (UBI), André Duarte (UBI), Miguel Camelo (IMEC), German Castellanos (ACC), Stephen Parker (ACC), Ultan Kelly (VIAVI), Burkhard Hensel (TUD), Christoph Sommer (TUD)
Date	28 June 2024
Related WP	WP2 | Network Digital Twin Modelling
Dissemination level	PU | Public, fully open

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

Document change history

Version	Date	Author	Description
V0.1	09/02/2024	Jean-Sébastien Sottet (LIST)	First document with a table of content.
V0.2	23/02/2024	All contributors	Initialisation of section content.
V0.3	20/03/204	Ayse Sayin (EBY), Ipek Arikan (EBY), Ramin Fuladi (EBY)	Proposition for initial security and privacy content.
V0.4	04/04/2024	German Castellanos (ACC)	Revision of telemetry framework, data collection approach.
V0.5	20/04/2024	Ayat Zaki Hindi (LIST)	Proposition for initial data model and standard analysis.
V0.6	10/06/2024	Jean-Sébastien Sottet (LIST), Sebastien Faye (LIST), Miguel Camelo (IMEC), German Castellano (ACC), Ana Pereira, Ultan Kelly (VIAVI)	Content harmonization and completion.
V0.7	13/06/2024	Miguel Camelo (IMEC), André Duarte (UBI)	Proposition for communication and deployment models and harmonization. Ready for peer-review.
V0.8	24/06/2024	Ayat Zaki Hindi (LIST), Jean-Sébastien Sottet (LIST)	Pre-final version after peer-review.
V1	28/06/2024	Sébastien Faye (LIST)	Final submitted version.

Quality control

	Name	Organisation	Date
Editor:	Jean-Sébastien Sottet	LIST	28/06/2024
Peer review 1:	Mario Franke	TUD	17/06/2024
Peer review 2:	Chris Murphy	VIAVI	17/06/2024
Authorised and submitted by (Project Coordinator):	Sébastien Faye	LIST	28/06/2024

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

Legal disclaimer

Executive Summary

This deliverable, D2.1 of the 6G-TWIN project, presents a comprehensive framework for data governance, privacy, and harmonization within the scope of Network Digital Twins (NDTs) in the context of 6G networks. The main objective of this document is to propose and define the necessary standards, architectures, and mechanisms required to manage data efficiently and securely, potentially in a federated environment – facilitating interoperability and compliance with regulations.

The main objectives of this deliverable are:

- To develop a standardized approach for integrating and managing data from diverse sources across future 6G network environments.
- To ensure compliance with data privacy regulations such as General Data Protection Regulation (GDPR) while enabling seamless data sharing and utilization.
- To create a scalable and interoperable data management framework that supports the integration of NDTs.

The proposed framework consists of several key components designed to guarantee a seamless interaction between the physical and virtual systems, as well as, between the digital system instances. We aim to set the baseline of a unified NDT data space, which is responsible for the management of data, and setting a taxonomy for data generated by the network, based on existing standards of previous generation.

To achieve that, we define the data points where data is collected from the physical network, including standardised interfaces in the context of 5G and O-RAN (Open Radio Access Network). We also focus on the telemetry framework that handles real-time data collection and exchange from and to the physical network. We ensure interoperability within this framework by harmonizing data representation and storage from different sources.

The document outlines the technical requirements for ensuring effective data management and security in a setup involving potentially multiple NDT instances:

- The need for a management control unit to coordinate data collection from the physical network.
- The creation of a distributed storage layer to manage data both centrally and across the federation.
- The implementation of telemetry framework interfaces and data information models to ensure interoperability.
- The establishment of a data harmonization and aggregation layer to clean and standardise data inputs.

A significant focus is placed on the governance, privacy, and security aspects of the data management framework. The deliverable identifies potential threats associated with integrating NDTs into the 6G architecture and proposes measures to mitigate these risks. Key considerations include:

- Ensuring compliance with GDPR and other relevant data protection regulations.
- Implementing robust access control mechanisms to protect sensitive data.
- Designing the system to be resilient against potential security breaches and data leaks.

The 6G-TWIN project aims to contribute significantly to the Smart Data Model initiative by proposing a new domain for network data. The project will also prototype a low Technology Readiness Level (TRL) data space connector, potentially serving as a reference

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

implementation for future network data spaces. The roadmap includes continuous development and refinement of the proposed data models and frameworks as the project progresses.

This deliverable sets the foundation for a harmonized, secure, and scalable approach to data governance in the evolving landscape of 6G networks, ensuring that data from various sources can be effectively utilized while maintaining strict compliance with privacy and security standards.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

Abbreviations and acronyms

Abbreviations and acronyms	
3GPP	The 3rd Generation Partnership Project
API	Application Programming Interface
CN	Core Network
CU	Centralized Unit
DN	Data Network
DRX	Discontinuous Reception
DT	Digital Twin
DU	Distributed Unit
EM	Element Manager
E-UTRA	Evolved Universal Terrestrial Radio Access
ETL	Extract, Transform, and Load
GA	Grant Agreement
GDPR	General Data Protection Regulation
gNB	Next Generation Node B
KPI	Key Performance Indicator
KVI	Key Value Indicator
LTE	Long Term Evolution
NDT	Network Digital Twin
NE	Network Element
NEF	Network Exposure Function
NR	New Radio
NRM	Network Resource Model
QoS	Quality of Service
RAM	Reference Architecture Model
RAN	Radio Access Network
RRM	Radio Resource Management
RSRP	Reference Signal Received Power
RSRQ	Reference Signal Received Quality
RU	Radio Unit
SDO	Standards Development Organization
SINR	Signal-to-Interference-plus-Noise Ratio
UAV	Unmanned Aerial Vehicle
UE	User Equipment
VNF	Virtualized Network Function (VNF)
VNFM	Virtual Network Function Manager (VNFM)
WP	Work Package

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

Table of Contents

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

Table of Figures

Table of Tables

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

1. Introduction

The rapid digitization of industries necessitates advancements in network technologies, particularly as we transition towards 6G systems. The 6G-TWIN project addresses this need by proposing an AI-native reference architecture that incorporates Network Digital Twins (NDTs).

In this chapter, we outline the objectives of 6G-TWIN, emphasizing its core mission to develop a sophisticated network framework, before specifically focusing on the key targets of this deliverable, presenting its structure and the contribution of the project's partners.

1.1. Aims and objectives

1.1.1. 6G-TWIN objectives

In response to the accelerating digitization across industries, the 6G-TWIN project emerges with a singular mission: to pioneer an AI-native reference architecture for the forthcoming 6G systems. At its core lies an ambitious vision to seamlessly integrate Network Digital Twins (NDTs) into the fabric of future networks, revolutionizing their optimization, management, and control in real-time.

To achieve its ambition, the 6G-TWIN has been built around several specific objectives:

- Specific Objective 1 (SO1) is central to the project's ambition, promising to design an open, federated and AI-native network architecture for the imminent 6G landscape. This architectural blueprint is designed to leverage NDTs, empowering intelligent data analytics and real-time decision-making, thereby laying the groundwork for unprecedented network efficiency and performance.
- Moreover, Specific Objective 2 (SO2) underscores the project's commitment to constructing a federated, graph-based NDT capable of accurately representing the intricate dynamics of highly dynamic and complex network scenarios. By establishing this digital sandbox for network planning, management, and control, 6G-TWIN paves the way for enhanced operational agility and adaptability.
- Simultaneously, Specific Objective 3 (SO3) drives the project's efforts towards implementing a robust modelling and simulation framework. This framework serves as a cornerstone for accurately portraying networked environments and rigorously testing the functionalities of the envisioned 6G architecture.
- Ultimately, as the culmination of its efforts, 6G-TWIN aims to materialize Specific Objective 4 (SO4) by testing, validating, and demonstrating the transferability of its solutions. Through the development of dynamic demonstrators catering to tele driving and energy efficiency use cases, the project aims to showcase the practical impact of its architectural foundation on real-world network scenarios, heralding a new era of connectivity and innovation.

Embedded within the core of the 6G-TWIN project lies a foundational framework driven by specific objectives aimed at revolutionizing the architecture of future 6G systems.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

1.1.2. Deliverable objectives

This document defines the data management mechanisms that support the deployment and operation of a NDT. It aims to define a standardized data model that allows the NDT system to interact with the physical network infrastructure in a consistent way, using the data collected from the different parts of the network. It also aims to propose appropriate data governance and to establish fair data exchange in the NDT context. More specifically, the objectives are:

- Define a coherent data space for NDTs, together with the requirements and information (i.e., data sources, knowledge, etc.) that this NDT should encompass.
- Define governance, storage, and physical-virtual access strategies to a well-defined data repository.
- Consider privacy and security mechanisms.
- Propose and validate data harmonization models to adapt data collected from heterogeneous domains and nodes to the defined data space.

1.2. Relation to other activities in the project

This deliverable, D2.1, is a crucial component of 6G-TWIN's Work Package 2 (WP2), which focuses on Network Digital Twin Modelling – as represented in the figure below. Its outcomes are essential for the successful execution of WP2 activities. Data is the cornerstone of the DT concept; specifically, for a NDT, it is imperative to thoroughly understand the various data types and their implications before developing any models or approaches. Consequently, this deliverable will serve as the foundation for upcoming WP2 tasks and deliverables, such as D2.2 (i.e., basic models) and D2.3 (i.e., functional models).

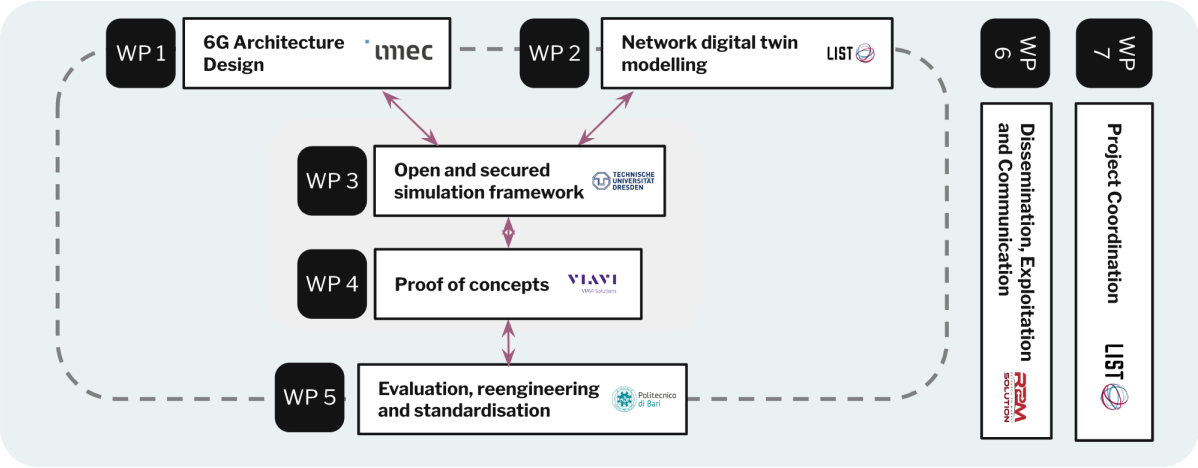

Figure 1. 6G-TWIN PERT chart.

The relationship with other tasks and work packages is also significant. D2.1 not only serves as a foundational document for WP2 but also integrates and supports the broader objectives of the project. It ensures that the data and preliminary models developed are coherent and robust enough to be built upon in subsequent deliverables and tasks, such as those in WP1 and WP3.

D2.1 is for instance connected to T1.2 (i.e., secured, scalable and distributed data exposure and collection for network monitoring). It also sets the stage for developing the low-TRL

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

prototypes to be used in WP3's open simulation framework. Finally, the work done in D2.1 feeds into T1.4 (i.e., federate management and orchestration of AI-based network functions and services) and informs the development of the simulation framework in WP3.

1.3. Document structure

This document covers various aspects related to data governance, privacy, and harmonization in the context of the 6G-TWIN project. Section 1 introduces the 6G-TWIN project by outlining its overarching goals, including the establishment of a robust framework for data governance and privacy in 6G networks. Additionally, the introduction connects the report to other activities within the project and acknowledges the contributions of various partners, emphasizing the collaborative nature of the initiative.

Section 2 introduces the concept of data spaces specifically designed for 6G Network Digital Twins (NDTs). It provides a background on data spaces and their importance for the NDT-enabled 6G networks, discussing key concepts and definitions. The architecture of O-RAN (Open Radio Access Network) is examined to provide a foundation for understanding subsequent sections. The section also proposes a taxonomy for an NDT data repository and outlines the necessary control plane data inventory required to construct a coherent data space.

Section 3 describes the technical requirements for data collection, distributed storage, and the telemetry framework. The challenges of handling heterogeneous data sources and the need for harmonization to ensure seamless integration across the 6G network are addressed. This section specifies the necessary technical prerequisites to ensure effective data management and utilization, highlighting the importance of advanced data collection mechanisms and a robust telemetry framework for real-time data acquisition.

Section 4 outlines the governance principles necessary for managing data within the 6G-TWIN framework. It discusses general governance principles and specific principles tailored to the project, emphasizing security measures to protect data and privacy mechanisms to ensure compliance with regulatory standards. This section is crucial for ensuring that data management strategies are not only technically sound but also legally and ethically robust.

Section 5 explores the integration of AI-native NDTs with the broader 6G infrastructure. It covers communication protocols to and from the physical network, as well as interactions with the simulation framework that supports decision-making processes. The section highlights how multiple digital twins (DTs) are connected, ensuring a cohesive and functional system architecture. It also discusses the role of AI in enhancing network functionalities, facilitating real-time analytics, and decision-making.

The conclusions in Section 6 summarize the findings and contributions of the report, reinforcing the importance of the proposed data governance, privacy, and harmonization strategies. This section reflects on the objectives outlined in the introduction, evaluates the success in achieving them, and discusses the implications for future work or further research needed to advance the 6G-TWIN project.

1.4. Contribution of partners

The following table present the contributions from all of the partners into the deliverable.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

Table 1. Partners contributions to the D2.1 deliverable.

Partner	Section(s)	Contributions
LIST	**1**, **2**, 4.1, **6**	Lead editor on the document; responsible for the content in Section 1 (Introduction), 2 (NDT data requirements), 4.1 (data governance) and 6 (conclusions).
IMEC	**5**	IMEC led the development and coordination of Section 5, giving more insights on the way physical and virtual networks can communicate and on NDT instantiation.
TUD	5.2	TUD explained the connection of the NDTs and their data to the simulation framework that is used to support higher-level decision making.
UBI	**3**, 5.3	Coordination of the work in Section 3 (data harmonization) and Section 5.3.
ACC	2.2.2, 3.2	Work on the O-RAN architecture and telemetry framework subsections.
EBY	**4**	Coordination of the work done in Section 4.
PX	2	Support in the definition of the data space in Section 2.
VIAVI	3.1, 5.1	Work on the data collection and distributed storage subsection.

Bold *numbers represent section technical leaders*

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

2. Towards 6G-NDT Data Spaces

2.1. Background on Data Spaces

A **data space** aims to provide the infrastructure and framework to support a data marketplace. A **data marketplace** is a data ecosystem that allows data providers to commercialize data as products for data consumers. It takes place both between actors of a given domain (e.g., energy, mobility, etc.) and between domains. A data space is not a standard open data repository. It focuses on data ownership and exchange between **data providers** and **consumers**. Therefore, the concept of data spaces covers different approaches that support the sharing of data between parties in a collaborative and fair way according to the European Data Strategy [1]. It acts as a regulated data market, governed by a data space governance authority.

Technically, a data space can be seen as a middleware that organizes the controlled sharing and use of data under a well-established governance framework. A data space [2] is both a multi-organizational agreement and a supporting technical infrastructure that enables data sharing between two or more participants. Participants in a data space may have a variety of pre-existing levels of trust. Some may have a prior relationship and trust each other, while others may have no relationship at all and be untrusted entities. Data spaces even allow data sharing between direct competitors. It provides answers to the technological challenges of data sharing:

- How do I find data?
- How do I publish my data securely?
- How do I share my data in a distributed environment?
- How do I maintain control of my data once it is shared?
- How do I ensure return on investment when sharing data?
- How do I ensure that data sharing is compliant with policies and regulations?

A data space also provides a business answer to these questions by offering a space to exchange data in a coordinated way, taking into account the value of the data. From this point of view, it can be seen as a data marketplace. As shown in Figure 2, a data space acts as a broker between two types of actors: data providers and data users/consumers. The data space should ensure the identity of each party and their common understanding of data (vocabulary in Figure 2). The data space can be seen as a store where data providers show their offerings and where data can be purchased by data users. Note that the value of the data is not necessarily expressed in terms of monetary terms, it can be other data that is returned or the result of the use of the data.

An important goal of a data space is to facilitate data sharing. In this context, data providers want to receive some return on their data assets. The 6G-TWIN project illustrates how this return does not have to be based on a purely financial transaction, but can also include the data producer benefiting from the value that data consumers can add and return to the data producer. This added value can include, for example, simulation services and planning decision-making for example.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

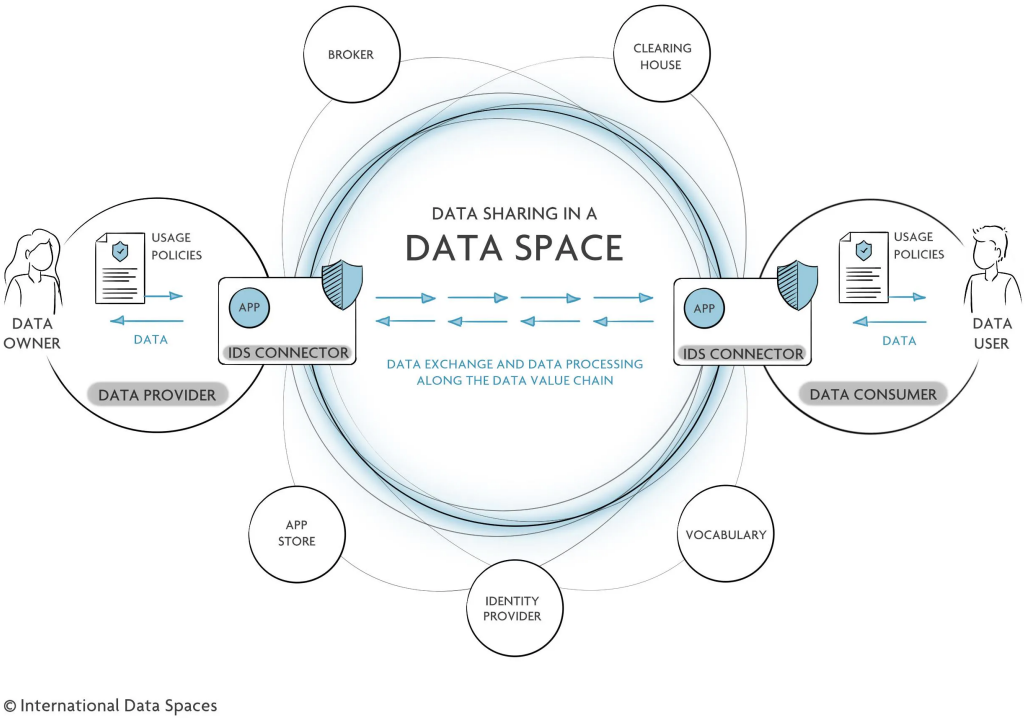

Figure 2. The concept of data space

Data connectors (IDS connector in Figure 2) ensure that more services are exposed (exposing offerings, ensuring identity, etc.). As a result, data-space data connectors are central entities that connect the existing system and its data sources to the data-space ecosystem. In the context of the International Data Spaces (IDS) initiative, their architecture and functionalities have been defined in the Reference Architecture Model (RAM) [3] .

Data connectors allow metadata to be added to the exchanged data. The metadata could conform to a standard representation to facilitate interoperability of data users (between providers and consumers). The standard representation may be a Smart Data Model. **Smart Data Models** [4] are structured data representations specifically designed to facilitate the development of solutions in various domains, including smart cities, smart agriculture, and smart industry. The Smart Data Models initiative aims to standardise the way data is represented, ensuring compatibility and ease of integration between different systems and services.

The use of a data connector ensures decentralised data storage and data integration is performed only through the connector. The data connector exposes the data provided through a catalogue (representing the metadata under a common, shared vocabulary) that represents the different data sets and offerings of the data providers. A data connector is then the technical software component that manages the data exchange through a data space protocol.

A **Data Space Protocol** [5], as defined by IDS, aims to facilitate data exchange between entities by defining the schemas and protocols required to publish data, negotiate agreements, and access data. It provides the fundamental specification to ensure interoperability for participant in data spaces (consumers and providers). Interoperability in this context also means agreeing on a level of trust at organisational and legal levels.

An **identity provider** offers the means to control the trust that both parties (providers and consumer) can have in each other during the data exchange. It helps to ensure the identity of the data participants and to define a trust protocol [6]. In particular, it could prevent data from

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

being made available to a competitor or to a data consumer outside a regulatory jurisdiction. Conversely, it also ensures a certain level of quality (e.g., through reputation) in the data provided. Once a level of trust has been established (including legal compliance), it is also important to ensure the respect of the contract, which acts as a common agreement between the consumer and the provider. In this context, the data space administrator (or data space authority) should also ensure that the contract respects the rules of the data space. The aspects related to the management of the data exchange and the role of the data space governance authority are discussed in more detail in Section 4.1.

As Digital Twins (DTs), and Network Digital Twins (NDTs) in particular, will rely heavily on data coming from different data sources owned by different partners, it is necessary to provide a data space approach. This approach will federate the collection and management of data used by the DT. It is recognised that there is a symbiotic relationship between DTs and data spaces [7]. In the first instance, a DT will aggregate the data provided by the various operators, and is therefore seen as a data consumer, and ultimately as a resource aggregator in relation to the domain(s) it covers. As a resource aggregator, it falls under the same governance status as any other data space actor. This will depend on the authority under which the DT is managed (e.g., network operator, public regulator, etc.). Secondly, it can also provide new data coming from its internal operations (e.g., synthetic data generated thanks to simulation scenarios), then it can be used as a data producer, in particular when participating in a multi-digital twin context (see Section 5.3).

2.2. NDTs for Mobile Networks: Key Concepts

As this project aims to study NDT aspects for 6G networks, we focus in this section on the characteristics of existing mobile networks as a baseline for future network evolution. NDTs are envisioned to support 6G systems by creating detailed virtual representations of the physical network's entities and processes, relying on models, real-time and historical data, and mapping between the physical and the virtual entities, as recommended by the ITU-T Y.3090 [8]. In the following sections, we introduce some basic definitions used to distinguish the different network entities and data, relying on the 3GPP vision.

While 3GPP is the lead Standards Developing Organisation (SDO) for defining the technical specifications of 5G, it works with various other organisations and SDOs. These groups contribute to different aspects of communication network technology, from network protocols and radio access to interoperability and industrial applications, ensuring a comprehensive and globally harmonised development and deployment of mobile networks.

2.2.1. Definitions and Concepts

Based on the definitions in 3GPP TS 28.622 [9], an **Information Object Class (IOC)** defines the management aspects of network resources by describing the information that can be exchanged via management interfaces in a technology-agnostic manner. A **network resource** can represent the intelligence, information, hardware, or software of a telecommunications network. IOCs have **attributes** that define various properties and can support operations for network management services and event notifications. IOCs provide a standardized approach to network resource management through the "class" structure.

Performance Management (based on 3GPP TS 32.401 [10]) aims to collect data from **Network Elements (NEs)** and **Virtualised Network Functions (VNFs),** to verify network

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

configurations, monitor traffic levels, assess resource access and availability, and ensure Quality of Service (QoS).

Performance measurements cover several aspects, including user and signalling traffic, network configuration effectiveness, resource access, and QoS parameters such as call setup delays and packet throughput. Additionally, performance data collection supports the detection and early resolution of potential network issues.

The administration of performance (6G-TWIN, 2024) measurements involves several stages: managing the measurement collection process, generating, and storing results, transferring data to Operational Support Systems (OSS), and presenting the results to network operators. The measurements scheduled by the **Element Manager (EM)** must be fully defined, including measurement types, resources, and granularity periods.

Measurement accuracy, reliability, and comparability across different vendor equipment are crucial to ensure consistent and interpretable data. Furthermore, measurement results need to reflect all occurrences of defined events and be attributable to the correct NEs and measurement types. The administration functions in the EM must handle numerous simultaneous measurement jobs, adapting to changes in network resources dynamically. Performance alarms are another vital aspect, where operator-defined thresholds trigger alarms upon being crossed, allowing for immediate attention to potential issues.

Main network components

Following the NDT architecture outlined in the D1.1 deliverable of 6G-TWIN (i.e., Architecture and Technical Foundations), the digital twin layer consists of basic and functional models. These models are based on the characteristics of different network entities, each modelled separately. To facilitate the modelling of various entities in a telecommunication network, we distinguish three main domains: User Equipment (UE), Radio Access Network (RAN), and Core Network (CN).

User Equipment (UE): This represents the connected devices in the network, such as mobile phones, connected Unmanned Aerial Vehicles (UAVs), and other terminals that are part of the NDT application.

Radio Access Network (RAN): Which is commonly characterised by the gNB entity (Next Generation Node B) providing RAN functionalities and connecting the users to the core network. In Long Term Evolution (LTE) (also known as 4G) and 5G, the radio interface is denoted by E-UTRA (Evolved Universal Terrestrial Radio Access) and NR (New Radio), respectively. Future generations will also include heterogeneous technologies such as satellites and Reconfigurable Intelligent Surfaces (RIS).

Core Network (CN): This includes Virtualised Network Functions (VNFs) responsible for collecting and managing data in the network and providing access to the Data Network (DN). DN can include operator services, Internet access, or 3rd party services. VNFs exchange data within CN and with RAN via standardised interfaces.

Figure 3 illustrates the basic 5G system architecture as described in 3GPP TS 23.501 [11]. It shows the three main components of the 5G system: UE, RAN, and CN, represented by the different standardised VNFs and interfaces.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

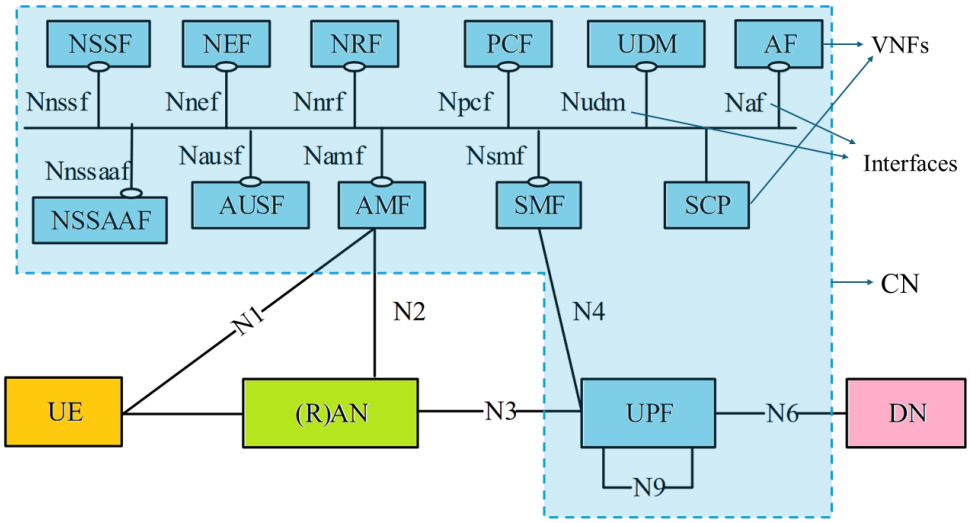

Figure 3. Main mobile network components, functions, and interfaces.

Exchanged data in mobile networks

Data within mobile networks is divided into user plane and control plane:

User Plane: Handles the actual data transfer between the UE and the network, including data streams such as web browsing, video streaming, and application data.

Control Plane: Manages signalling required to establish and maintain connections, including session management, mobility management, and connection setup.

Both user and control plane data are useful for building an NDT. For example, user plane data aids in developing user behaviour models and optimizing network data caching and content delivery. Control plane data is crucial for obtaining real-time and historical performance indicators, optimizing network attributes, and detecting anomalies.

In this document, we focus on control plane data to characterise the different network entities and develop a taxonomy for NDTs.

2.2.2. 3GPP and O-RAN Architectures

To clarify some of the descriptions contained in the following sections, it is useful to provide a brief overview of what could be the 6G network architecture, highlighting the key RAN components and the interfaces between them. It is important to note that it is not yet clear what exactly 6G will entail, so some of what is presented here is necessarily, somewhat speculative. Figure 4 below, shows the vision of the Open 6G logical architecture, which has evolved from a 5G base architecture. The base 5G architecture is defined by 3GPP and this was further disaggregated by the O-RAN Alliance [12] with additional standard interoperable interfaces with a view to promoting ‘open’ multi-vendor solutions, stimulating competition and innovation, increasing supplier diversity, reducing cost, and bringing intelligence to the RAN via the RAN Intelligent Controller (RIC).

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

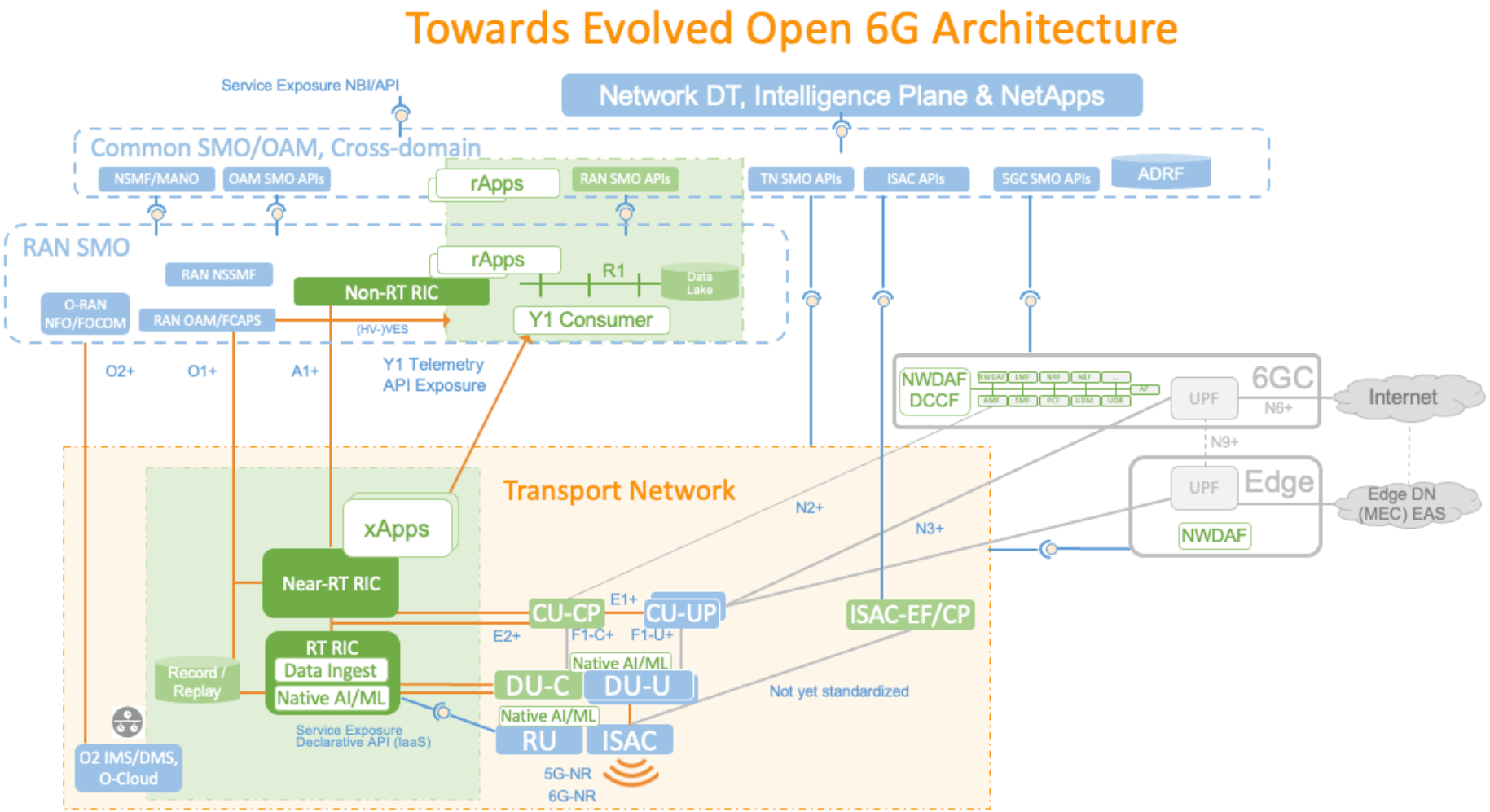

Figure 4. Towards a 6G Logical Architecture.

Some interfaces are labelled with a "+" symbol. This is to indicate that these interfaces will be assumed evolutions of existing 5G interfaces.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

Figure 4 shows the main building blocks that can reasonably be expected to constitute a 6G network; The Radio Access Network (RAN), the 6G Core Network (6GC), a Service Management and Orchestration (SMO) layer and potentially an Edge Computing component to handle Integrated Sensing and Communication (ISAC) functionality. The focus of this section is on RAN telemetry so the SMO, 6GC and ISAC components are not further explained here. It is sufficient to say that the SMO is responsible for managing the lifecycle of network services, the 6GC provides the essential functions required for managing connectivity, data routing, and service delivery. In addition, the ISAC is a concept that merges the functionalities of wireless communication and sensing into a single system, allowing physical network information to be gathered into the digital network while maintaining efficient communication capabilities.

The 3GPP and O-RAN specifications allows for splitting the RAN functionality into four logical components:

1. **The Radio Unit (RU),** provisioned with RF circuitry.
2. **The Distributed Unit (gNB-DU or DU),** hosting gNB real-time functions.
3. **The Central Unit (gNB-CU or CU),** hosting gNB non-real-time functions.
4. **The RAN Intelligent Controller (RIC),** providing intelligent control capabilities.

An overview of the functionality of these components is provided below.

The RU

The RU or O-RAN Radio Unit O-RU is the unit at the edge of the access network. It interfaces with the O-DU northbound via standardized O-RAN Interfaces and with the 5G/6G UE via the air Interface. All the O-RU interfaces are following standards based on:

- O-RAN Interface between the O-RU and O-DU over 10Gbps Ethernet links.
- Orthogonal Frequency-Division Multiple Access (OFDMA) Air Interface between the O-RU and the User Equipment according to the 3GPP standard (Sub 6GHz band).

The DU

The DU is usually located in a range of several kilometres from the RU and runs parts of the 3GPP PHY layer, MAC and Radio Link Control (RLC) layers. This logical node includes a subset of the gNB functions, depending on the functional split option, and its operation is controlled by the CU, and possibly by the RIC. A single DU controls one or more RUs via a Fronthaul Interface (FHI).

The CU

The Central Unit (CU) implements the ‘Layer 3’ part of the NG-RAN gNB. The CU is split between the control plane and user plane:

- CU-CP: Control-plane of CU, implementing the RRC functions as described in [13] chapter 5.
- CU-UP: User plane function of CU, implementing the Packet Data Convergence Protocol (PDCP) and the Service Data Adaptation Protocol (SDAP) (mapping of PDU sessions and their QoS flows to the access stratum Data Radio Bearers (DRB)).

The RAN Intelligent Controller (RIC)

The O-RAN architecture introduces a new software-based network element called the RAN Intelligent Controller (RIC). The RIC is designed to monitor, control, and optimise the RAN network operation. It is involved in controlling and optimising aspects such as Radio Resource

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

Management (RRM), Power Management, Interference Management and Mobility Management. The RIC provides a set of Application Programming Interface (APIs) to developers to enable them to build RAN intelligence in the form of applications called xApps or rApps.

The RIC consists of two components.

1. The non-real-time RAN Intelligent Controller (non-RT RIC): Which manages events and resources with a response time of one second or more. rApps is the term for applications that are developed to control and manage the RAN.
2. The near-real-time RAN Intelligent Controller (near-RT RIC) is a component that manages events and resources within the RAN with a fast response time of up to 10 milliseconds. This ensures efficient and quick adjustments to the network. The applications designed to control and manage the RAN through this controller are known as xApps. These xApps use the capabilities of the near-RT RIC to improve network performance and adaptability.

Due to the indicated response times, the near-RT RIC would typically be deployed on the network edge. As it has less stringent response times, the non-RT RIC can either be deployed centrally or on the network edge.

Principal 3GPP and O-RAN interfaces

In the following sections, we describe the principal 3GPP and O-RAN interfaces that are referred to later in this document.

3GPP F1 Interface: The F1 interface [14] is the interface that sits between the CU and the DU. It is further divided into the F1-C (F1 Control) and F1-U (F1 User Plane) sub-interfaces.

3GPP E1 Interface: The E1 interface [15] sits between the CU-CP and CU-UP. It is a control interface and is responsible for interface configuration, setup, reset and error reporting.

O-RAN E2 Interface: The E2 interface [16] links the RIC and the CU and the RIC and the DU. When referring to the E2 interface the CU and DU are called E2 Nodes. This interface allows the RIC to control the behaviour of the E2 nodes also to collect RAN metrics from them. These metrics can either be triggered or sent periodically.

O-RAN O1 Interface: The O1 interface [17] is the interface between the Service Management and Orchestration (SMO) and O-RAN managed network elements i.e., it is the interface between the SMO and the CU, the SMO and the DU and the SMO and the near-RT RIC. It is a management interface, and it covers aspects such as element discovery, registration, provisioning and monitoring.

O-RAN O2 Interface: The O2 interface [18] is how the SMO communicates with the “Cloud” that it resides in. It is used by the network operator to (re)configure and deploy network elements and upgrade the system.

O-RAN A1 Interface: The A1 interface [19] links the non-RT RIC and the near-RT RIC and is purpose is to support policy management, service enrichment and machine learning. It provides a non-real-time control loop towards the near-RT RIC used for example, for sharing policy-based guidance with the near-RT RIC e.g., for setting near-RT RIC high-level optimisation goals.

O-RAN Open Fronthaul Interface: The Open FH (Fronthaul) Interface [20] links the DU and RU. It includes the CUS (Control User Synchronization) Plane and M (Management) Plane.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

2.3. Taxonomy for NDT Data Repository

This document aims to provide a preliminary framework for the data collection needed to build NDT models for 6G networks. It provides a draft for the so-called “vocabulary” in the data space framework illustrated in Figure 2. The proposed approach focuses on data generated directly from the physical network and does not comprehensively include contextual data such as application requirements, building environments, sensors, and other critical factors necessary for creating accurate NDT models. These contextual elements can pertain to different domains that are already well-defined by other standards and consortiums and can be plugged into our NDT, e.g., smart cities data models [4].

2.3.1. Control Plane Data

Control plane is responsible for exchanging data critical to network operation and management. This data includes network attributes and performance measurements.

Network Attributes

These include network topology, configuration aspects, and data exchanged to control the performance of network elements (NEs). Key attributes include:

1. **Radio Resource Management (RRM) Parameters**: Controls allocation and management of radio resources to ensure efficient usage, including power control, handover decisions, and admission control.
2. **Quality of Service (QoS) Parameters**: Ensures different types of traffic are treated according to their priority and service requirements by setting up bearers with specific QoS profiles.
3. **Security Configurations**: Manages authentication, authorization, and encryption processes to secure communication between devices and the network.
4. **Network Slicing**: Creates multiple virtual networks over a single physical infrastructure, each optimized for specific use cases (e.g., enhanced mobile broadband, massive IoT, ultra-reliable low latency communication).

Performance Measurements

These measurements capture network behaviour based on attribute configuration and are communicated between different NEs. Key performance measurements include:

1. **Signal Quality Metrics**: Measurements like Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and Signal-to-Interference-plus-Noise Ratio (SINR) assess link quality and inform decisions about handovers and power control.
2. **Traffic Load**: Monitors load on network elements to ensure balanced resource allocation and prevent congestion.
3. **Mobility metrics**: Tracks location and movement patterns of user devices to manage handovers and ensure seamless connectivity.
4. **Network Performance Metrics**: Collects data on Key Performance Indicators (KPIs) such as latency, throughput, and packet loss to optimize network performance and identify areas for improvement.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

5. **Energy Consumption**: Monitors power usage of network elements to improve energy efficiency and manage operational costs.

2.3.2. NDT data inventory

As previously mentioned, the network comprises three primary elements to be twinned: UE, RAN, and CN. For RAN and CN, we review the most important IOCs and their related attributes and measurements based on the latest 3GPP specifications. In this context of 6G-TWIN, we rely on TeraVM RAN Scenario Generator (RSG) and TeraVM Core Emulator as means to provide the necessary NDT datasets, as these tools cover most of the aspects of mobile RAN and CN, in compliance with standards. These tools are further detailed in section 5 of this document.

For UE, we focus on the measurements that can be performed by them rather than their attributes, as configuring UEs is done via command messages that are not directly related to specific attributes. Configuring UEs by the network is performed via the Non-Access Stratum (NAS) and New Generation Application Protocol (NGAP), which connects the UE and CN, described in 3GPP TS24.501 [21] and TS38.413 [22], respectively. Contrary to RAN and CN measurements, UE measurements are not directly transferred to the network; instead, a specific measurement triggering and reporting procedure is followed, detailed in 3GPP TS 38.331 [23]. The network configures the UE with measurement parameters which triggers the reporting procedure based on predefined events. In the RRC_CONNECTED state, the UE actively communicates with the network, managing data transfer, Discontinuous Reception (DRX) cycles, carrier aggregation, and dual connectivity. The UE capabilities include monitoring paging and control channels, providing channel quality feedback, and performing neighbouring cell measurements.

In the following tables, we connect each RAN IOC, CN IOC, and UE measurement category to our proposed use-cases in 6G-TWIN, indicating the relevance of each IOC on a scale from 1 to 4, with 1 being the least relevant and 4 being the most relevant. We note that the following tables are not exhaustive, and they are heavily influenced by the future work to be done in the different work packages of 6G-TWIN.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

Table 2. Radio Access Network (RAN) Data Inventory

Information Object Class (IOC)	Description	Attributes (3GPP TS 28.541)	Measurements (3GPP TS 28.552)	Relevance to 6G-TWIN's use-case 1	Relevance to 6G-TWIN's use-case 2
NRCellCU	NR Centralized Unit is responsible for the management of inter-cell mobility and neighbour relations via automatic neighbour relation.	cellLocalId; pLMNInfoList; Attribute related to role; nRFrequencyRef;	RRC connection number RRC connection establishment and Re-establishment UE-associated logical NG-connection DRB (Data Radio Bearer) related measurements PDU Session Management Measurements related to MRO (HO) 5QI x QoS Flow Duration (valid only for split-gNB mode) Packet Loss Rate, Packet Drop Rate, Packet Delay, IP Latency	1	1
NRCellDU	NR Distributed Unit provides performance statistics based on the configured sector, carrier frequencies, and channel bandwidth. The measurements are performed in downlink.	cellLocalId; operationalState; administrativeState; cellState; pLMNInfoList; nRPCI; nRTAC; arfcnDL; arfcnUL; arfcnSUL; bSChannelBwDL; ssbFrequency; ssbPeriodicity; ssbSubCarrierSpacing; ssbOffset; ssbDuration; bSChannelBwUL; bSChannelBwSUL;	Packet Delay, Radio resource utilization UE throughput Number of Active UEs CQI (Channel Quality Indicator) related measurements Transmit power utilization measurements Received Random Access Preambles PHR (power headroom) Measurement	4	4
NRSectorCarrier	NR Sector Carrier can specify the transmission direction: downlink, uplink, or both. This IOC represents the resources of each transmission point associated to corresponding cell(s), such as, physical antenna location, frequency, and bandwidth.	txDirection; configuredMaxTxPower; configuredMaxTxEIRP; arfcnDL; arfcnUL; bSChannelBwDL; bSChannelBwUL;	-	1	4
BWP	The BandWidth Part (BWP) is related to downlink, uplink or supplementary uplink resource grids, including subcarrier spacing, cyclic prefix, and location.	bwpContext; isInitialBwp; subCarrierSpacing; cyclicPrefix; startRB; numberOfRBs;	-	1	2

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

NRCellRelation	NR Cell Relation represents a neighbour cell relation from a source NRCellCU instance to a target NRCellCU instance. Neighbour cell relations are unidirectional.	nRTCI; cellIndividualOffset; isRemoveAllowed; isHOAllowed; isESCoveredBy; isENDCAllowed;	-	1	2
NRFreqRelation	NR Frequency Relation, together with the target NRFrequency, represents the frequency properties applicable to the referencing NRCellRelation.	offsetMO; blackListEntry; blackListEntryIdleMode; cellReselectionPriority; cellReselectionSubPriority; pMax; qOffsetFreq; qQualMin; qRxLevMin; threshXHighP; threshXHighQ; threshXLowP; threshXLowQ; tReselectionNr; tReselectionNRSfHigh; tReselectionNRSfMedium;	-	1	1
S-NSSAI	Network Slice Selection Assistance Information (NSSAI) includes SST (Slice/Service type) and optional SD (Slice Differentiator).	sST; sD;	-	2	1
NRFrequency	This IOC represents certain NR frequency properties.	absoluteFrequencySSB; sSBSubCarrierSpacing; multiFrequencyBandListNR;	-	2	2
CommonBeamformingFunction	The associated attributes to this IOC configure the wanted coverage area and radiation pattern on a sector carrier.	coverageShape; digitalTilt; digitalAzimuth;	-	2	4
Beam	This IOC represents the per-Beam information which can be used for instance in beam performance management or troubleshooting performance problems. Measurements are generated in the RAN.	beamIndex; beamType; beamAzimuth; beamTilt; beamHorizWidth; beamVertWidth;	Intra-NRCell SSB Beam switch Measurement RSRP Measurement SSB beam related Measurement	2	4
RRMPolicyRatio	Radio Resource Management (RRM) policy ratio.	rRMPolicyMaxRatio; rRMPolicyMinRatio; rRMPolicyDedicatedRatio;	-	2	4
RRMPolicy_	RRM policy properties must be subclassed to be instantiated. The attributes define respectively the type of resource (PRB, RRC connected users, DRB usage etc.) and the RRM policy members related to this policy.	resourceType; rRMPolicyMemberList;	-	1	4
RimRSGlobal	Remote Interference Management (RIM) Reference Signal (RS) represents the global/common resource allocated for the whole network. The associated	frequencyDomainPara; sequenceDomainPara; timeDomainPara;	-	2	2

Co-funded by the European Union
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.
6GSNS

	configured parameters must be applied to all sets of RIM RS Resource across gNBs/cells in the network.				

Table 3. Core Network (CN) Data Inventory

Information Object Class (IOC)	Description	Attributes (3GPP TS 28.541)	Measurements (3GPP TS 28.552)	Relevance to 6G-TWIN's use-case 1	Relevance to 6G-TWIN's use-case 2
AMFFunction	Access and Mobility Management Function (AMF)	pLMNIdList; aMFIdentifier; sBIFQDN; sNSSAIList; managedNFProfile; commModelList;	Registered subscribers measurement Registration procedure related measurements Mobility related measurements UE Configuration Update procedure related measurements Authentication procedure related measurements Measurements related to registration via trusted/untrusted non-3GPP access Measurements related to Service Requests via Trusted/Untrusted non-3GPP Access	4	1
SMFFunction	Session Management Function (SMF)	pLMNIdList; nRTAClist; sBIFQDN; sNSSAIList; managedNFProfile; commModelList;	Session management QoS flow monitoring Performance measurement for N4 interface	2	1
UPFFunction	User Plane Function (UPF)	pLMNIdList; nRTAClist; sNSSAIList; managedNFProfile; supportedBMOList;	Interfaces related measurements GPRS Tunnelling Protocol (GTP) packets delay in UPF One way packet delay between NG-RAN and PSA UPF Round-trip packet delay between PSA UPF and NG-RAN One way packet delay between PSA UPF and UE QoS flow related measurements	1	3
N3IWFFunction	Non-3GPP Interworking Function (N3IWF) interface	pLMNIdList; commModelList;	PDU Session Resource management QoS flow management	1	1
PCFFunction	Policy Control Function (PCF)	pLMNIdList; sBIFQDN; sNSSAIList; managedNFProfile;	AM policy association/authorization related measurements	1	1

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

		commModelList; supportedBMOList;	SM policy association/authorization related measurements UE policy association related measurements Event exposure related measurements		
UDMFunction	Unified Data Management (UDM)	pLMNIdList; sBIFQDN; sNSSAIList; managedNFProfile; commModelList;	Mean and maximum number of registered subscribers through UDM Mean and maximum number of unregistered subscribers through UDM Subscriber data management related measurements Parameter provisioning related measurements	2	2
UDRFunction	Unified Data Repository (UDR)	pLMNIdList; sBIFQDN; sNSSAIList; managedNFProfile;	Data management related measurements	3	3
NRFFunction	Network function Repository Function (NRF)	pLMNIdList; sBIFQDN; sNSSAIList; nFProfileList; cNSIIdList;	NF service registration, update, and discovery related measurements	1	1
LMFFunction	Location management function (LMF)	-	Location determination, notification, and context transfer related measurements	4	1
NSSFFunction	Network Slice Selection Function (NSSF)	pLMNIdList; sBIFQDN; sNSSAIList; cNSIIdList; managedNFProfile; commModelList;	Network slice selection and S-NSSAI (Network Slice Selection Assistance Information) availability related measurements	2	1
NWDAFFunction	Network Data Analytics Function (NWDAF)	-	Measurements related to the NWDAF analytics service, data collection, service provisioning, and others.	3	3
NEFFunction	Network Exposure Function (NEF)	-	Measurements related to application triggering, PFD management, AF traffic influence, and others.	3	3
FiveQICharacteristics<<dataType>>	This data type specifies the 5QI value and the corresponding QoS characteristics for a 5QI.	fiveQIValue; resourceType; priorityLevel; packetDelayBudget; packetErrorRate; averagingWindow; maximumDataBurstVolume;	-	1	1

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

QoSData <<dataType>>	This data type specifies the QoS control policy data for a service flow of a PCC rule.	qosId; fiveQIValue; maxbrUl; maxbrDl; gbrUl; gbrDl; arp; qosNotificationControl; reflectiveQos; sharingKeyDl; sharingKeyUl; maxPacketLossRateDl; maxPacketLossRateUl; extMaxDataBurstVol;	-	1	1
ARP <<dataType>>	This data type specifies the allocation and retention priority of a QoS control policy.	priorityLevel; preemptCap; preemptVuln;	-	1	1

Table 4. User Equipment (UE) Data Inventory

Category	Data name (3GPP TS 38.215)	Data description	Relevance to 6G-TWIN's use-case 1	Relevance to 6G-TWIN's use-case 2
Synchronization Signal (SS)	SS reference signal received power (SS-RSRP) SS reference signal received quality (SS-RSRQ) SS signal-to-noise and interference ratio (SS-SINR) SS reference signal received power per branch (SS-RSRPB) SS reference signal antenna relative phase (SS-RSARP)	Synchronization signal (SS) is used to synchronize UE with a gNB. These signals consist of the Primary Synchronization Signal (PSS) and the Secondary Synchronization Signal (SSS).	1	1
Channel State Information (CSI)	CSI reference signal received power (CSI-RSRP) CSI reference signal received quality (CSI-RSRQ) CSI signal-to-noise and interference ratio (CSI-SINR)	Channel State Information (CSI) reference signals are used by the UE to estimate the channel and report channel quality information to the gNB.	4	4
Global navigation satellite system (GNSS)	UE GNSS Timing of Cell Frames for UE positioning for E-UTRA UE GNSS code measurements UE GNSS carrier phase measurements	UE location-related measurements via GNSS.	4	1
Wireless Local-Area Network (WLAN)	IEEE 802.11 WLAN RSSI	Received Signal Strength Indicator (RSSI) from IEEE 802.11 WLAN.	1	1
Evolved Universal Terrestrial Radio Access (E-UTRA)	Reference signal time difference (RSTD) for E-UTRA System Frame Number (SFN) and frame timing difference (SFTD) E-UTRA RSRP / E-UTRA RSRQ / E-UTRA RS-SINR	Measurements related to E-UTRA, used by the UE for synchronisation and channel quality estimation.	1	2

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

Sounding Reference Signal (SRS)	SRS reference signal received power (SRS-RSRP)	SRS transmission occurs periodically at appropriate power levels, such that a gNB can measure the quality of its synchronization to the served UE.	1	1
Cross Link Interference (CLI)	CLI Received signal strength indicator (CLI-RSSI)	Measures average interference over the configured measurement bandwidth from all sources, including co-channel serving and non-serving cells, adjacent channel interference, and thermal noise.	2	2
Sidelink measurements	Physical sidelink broadcast channel reference signal received power (PSBCH-RSRP) Physical sidelink shared channel reference signal received power (PSSCH-RSRP) Physical sidelink control channel reference signal received power (PSCCH-RSRP) Sidelink received signal strength indicator (SL RSSI) Sidelink channel occupancy ratio (SL CR) Sidelink channel busy ratio (SL CBR)	Measurements related to UE sidelink communication.	3	1
Downlink (DL)-related measurements	DL PRS reference signal received power (DL PRS-RSRP) DL reference signal time difference (DL RSTD) UE Rx – Tx time difference	Measurements related to UE positioning, mainly through DL Positioning Reference Signals (PRS) and time difference.	3	1

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

2.4. 6G-TWIN Data Space Proposal

The 6G-TWIN data space will be organised to support the collection and the fair sharing of network data with the purpose of building and operating a digital twin.

It is composed of the following components:

- Data connector that will ensure all the connections between the different source of data and consumers of data respecting the defined data governance.
- Data governance that defines rules and policy for preserving data privacy, security, sovereignty, and the fairness of data exchanges.
- Data collection that ensuring the capturing, clearing, and storing of data (with regard the data governance rules).
- Communication (API) from and to the Digital twin.

The semantic harmonization of data is an important topic in the context of Data Space. The different stakeholders should use the same language. Based on the previous section taxonomy (see Section 2.3), the 6G-TWIN project will propose a new smart-data model [4] dedicated to Network. **As of June 2024, a preliminary submission of this model has been incubated on the smart data model GitHub repository[1] [24]**.

The proposed data space positions the NDT not only as data consumer but also potentially as data provider. By using the holistic view built by the NDT, it provides data owner with a comprehensive insight and enhanced decision-making capabilities for its own business. . In this context, one digital twin can act as producer of data (e.g., synthetic data obtained from simulations). It will also manage the connection with third party applications that could use the NDT data.

The defined Data space will also act as an intermediary communication media amongst digital twins (DT as provider and consumer) in the context of Federated Digital twins. Each digital twin of the federation should be able to support a distributed data storage that respects the data governance rules.

Figure 5 is showing the overall data organisation and data flow around the central concept of data space. It also shows how the reminder of the deliverable is organised. The Sections 3.1 is depicting the telemetry and control framework. Section 3.2 is highlighting the harmonized, whilst section 3.3 is about data collection and distributed storage. Sections 4.1 to 4.3 are presenting respectively the general governance of our data space proposal. We focus, in our specific context, on the two main pain points: data privacy and data security. Finally, section 5.1 is presenting communication aspects of data from and to the data space. The section 5.2 is dedicated to internal communication (API) inside the NDT, between the NDT data and the

[1] https://github.com/smart-data-models/incubated/tree/master/CROSSSECTOR/6G-TWIN

simulation. Section 5.3 is providing insights on how to operate a federated NDT notably through data exchange that can be regulated by a Data space.

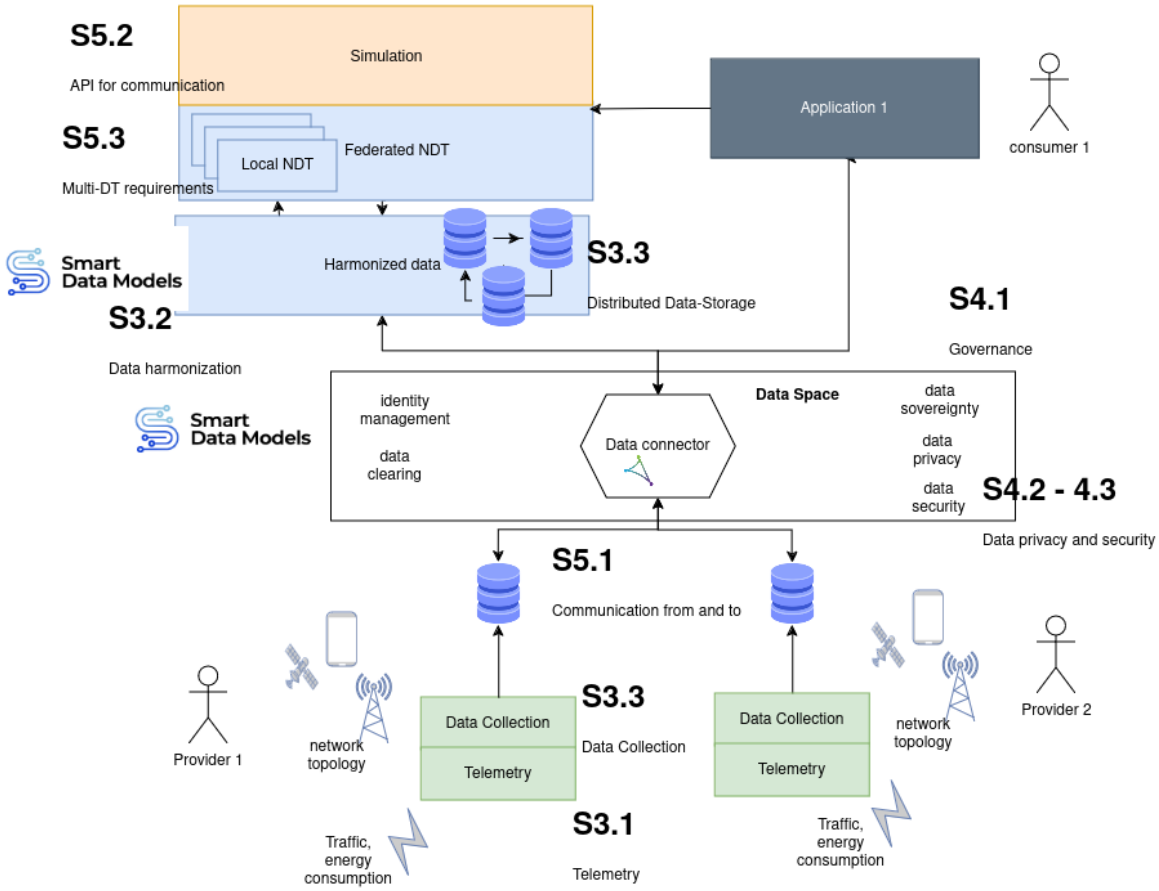

Figure 5. 6G-TWIN Data Management - Data space for building NDT

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

3. Data Flow Requirements

The concept of data management for 6G NDTs is crucial to unlocking and understanding the full potential of the new iteration of networking. In this sense, having a real-time replica of the physical network (including its assets and connections) is of utmost importance to unlock better network management, control, and monitoring. For instance, gathering the network's data in real-time and having it on an NDT allows for more precise management and more straightforward diagnosis of problems as well as what-if analysis, simulation scenarios and optimization. Effective data management is essential for ensuring these digital twins' accuracy, efficiency, and security.

The concept of data management for 6G NDTs unfolds into a complex system of managing heterogeneous data sources and data coming from different sources such as sensors, network elements, and applications. All these diverse data sources converge in a point of storage, which can be centralised and distributed. This central NDT acts as a data repository within the data domain and is responsible for collecting, storing, and managing data to build accurate and up-to-date models in the NDT. As a secondary effect, this data repository is also important in the context of federation, a topic that will be further discussed in Section 5 of this document.

At this stage, the data flow that the 6G-TWIN project aims for is becoming apparent. More specifically, we will push forward reference architectures and propose a central data flow to create a unified vision of this space. All of this will reach a common data space (better explained in Section 2) where all relevant datasets will live, and applications shall be provided to pave the way for the 6G KPIs/KVIs to be reached.

Of course, the concept of data management for 6G NDTs is not without its challenges. These challenges, which must be addressed in the context of the project, are related to the data flow for an instance of an NDT. They include:

1) **Data Volume and Velocity**: The sheer volume and velocity of data generated by 6G networks require efficient data management strategies to ensure timely processing and analysis. In this sense, the NDT instance shall be prepared to receive and process data in different manners, namely: (1) of high volume, which includes data points, e.g., datasets, with a high amount of data (terabytes or more); and (2) of high velocity, which includes incoming data streams from the network in (near) real-time. These aspects showcase how the architecture and data flow shall be ready for both types of processing: batch and real-time.
2) **Data Variety**: The anticipated diverse nature of data sources and formats in 6G networks demands flexible data management systems that handle various data types and structures. Here, the primary purpose shall be to aggregate, clean and transform the data into a common standard and common data format. Such processes shall represent the core of the data flow. They shall be the central harmonization component in each NDT (that can later be federated as per Chapter 5 descriptions).
3) **Data Integration and Interoperability**: A standard reference API shall harmonize a more integrated scenario after harmonizing in a standard data model. Integrating data from various sources and ensuring interoperability between systems and applications are essential for seamless data management in 6G NDTs.
4) **Data Security and Integrity**: Ensuring the security and integrity of data in 6G NDTs is critical, as any breaches or tampering can compromise the accuracy and reliability of the digital twin. Security and integrity requirements appear as overarching

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

requirements for all of the 6G NDT specifications and shall be present in all of the decisions on design and implementation.

With all of this in mind, the 6G-TWIN project is at the forefront of innovation, revolutionising how we handle massive amounts of data from diverse sources in real-time, and creating the NDT. This section delves into the integrated system that forms the core of our data management, focusing on the building blocks that are set to support and enhance 6G networks, as outlined below.

Telemetry Framework for 6G NDT

The telemetry framework in a 6G network is designed to handle real-time data collection and transmission with minimal latency. This framework collects telemetry data from devices, network nodes, and applications, ensuring continuous monitoring and feedback. The framework shall support multiple communication protocols and secure data transmission, providing robust and reliable data flow across the network. In 6G-TWIN, the collection framework will ideally be deployed in each network node. It will communicate the necessary information to the central digital twin, enabling a federation of all nodes in a single NDT.

Heterogeneous Data Sources and Harmonization for 6G NDT

In the scope of 6G-TWIN, the NDT shall integrate data from several heterogeneous sources, including different types of sensors, devices, and communication protocols. Here, a component to harmonise the information is of utmost importance and will be critical for the NDT. Moreover, harmonisation will enable data analytics and support complex 6G applications like enhanced AI-driven services.

Data Collection and Distributed Storage for 6G NDT

In the 6G ecosystem, data collection and distributed storage are fundamental to managing the enormous volumes of data generated by billions of connected devices. This component captures data from sensors, IoT devices, user equipment, and network infrastructure. Using a distributed storage model, NDTs can ensure data is stored across multiple locations, enhancing fault tolerance, scalability, and data access speeds. One of the most critical aspects of this is the place where the data sits and by whom it can be accessed. In this sense, 6G-TWIN will push a distributed model, in which each dataset owner will keep their data to themselves but allow access to the NDT to use the necessary information. Access policies shall mediate all of this.

In Figure 6 below, we outline the high-level data flow that will be the main connection point between the different components in the NDT architecture. With this in mind, we are making the following assumptions:

- **Telemetry Framework Layer**: This layer is the provider of information and sees every data point and dataset, regardless of their nature (public, private, protected). Here the data is generated and can be gathered by other components in the NDT.
- **Heterogeneous Data Sources and Harmonisation Layer:** This layer will be the central component of the ingestion infrastructure, since it holds the components responsible for doing all the harmonisation, cleaning and aggregation. This layer will deliver all of the data points represented in Smart Data Models as defined in section 2. In this layer (which is better explained in Chapter 3.2) the Data Aggregation component is responsible for collecting or receiving the data coming from the Telemetry Framework components. Then, it cleans it, transforms it and delivers it to the next layer.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

- **Data Collection and Distributed Storage Layer**: This layer shall be the data storage to support the NDT. Data on this layer should be accessed in two manners: (1) Distributed, where data sits distributed on the network and the Identity and Access Control mechanisms of the NDT have access policies to each of the data points. This represents each of the nodes on the federation; (2) Centrally, where each of the federation members choose which data to share with the central node. However, the central node holds the responsibility of defining the minimum data that needs to be shared to implement its requirements.

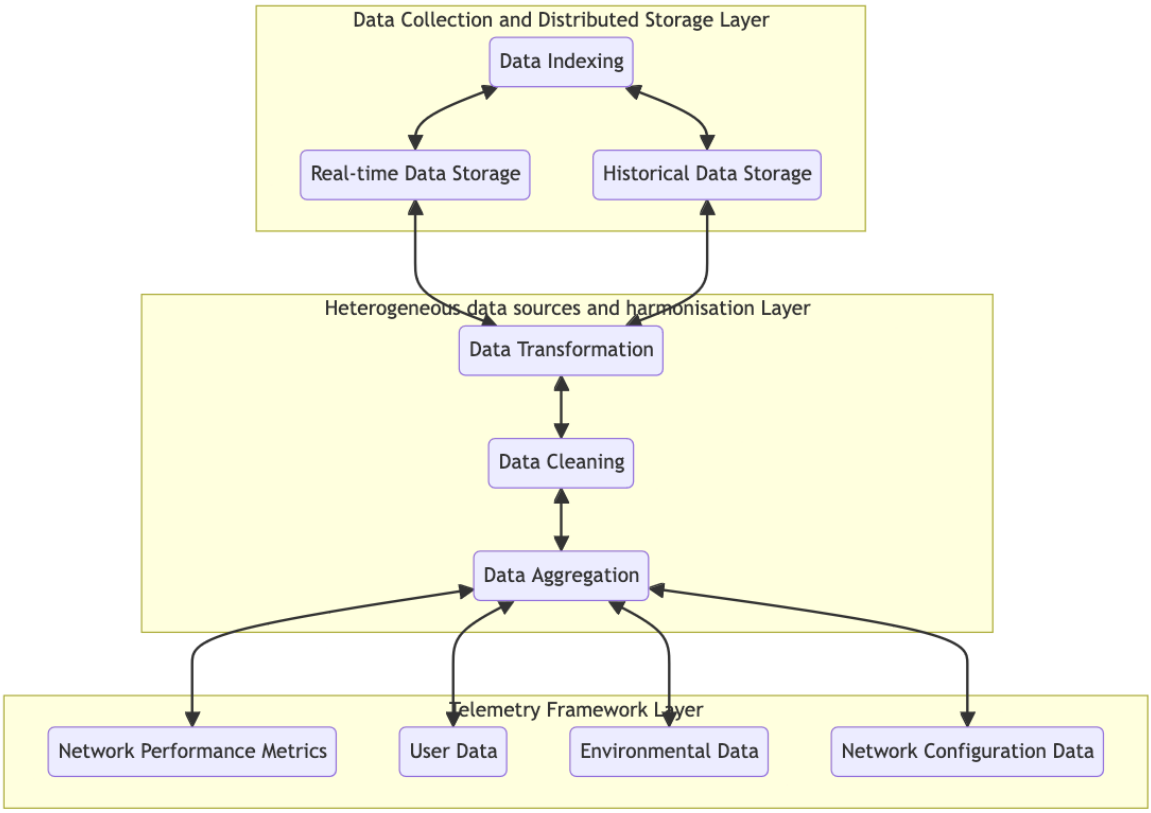

Figure 6. High-Level Data Flow Diagram

3.1. Telemetry and Control Framework

The Telemetry and Control Framework is developed to streamline and organize telemetry data and provide translator control capabilities within O-RAN technology environments. This framework is particularly crucial for handling data from various segments of O-RAN infrastructure and is structured around several key components.

At the heart of the system is the telemetry gateway (TGW), which plays a pivotal role in ensuring interoperability across different parts of the O-RAN setup. Given that O-RAN interfaces such as E2, O1, F1 or A1 might not always be fully designed or O-RAN-compliant, the TGW steps in to bridge these gaps. It translates or regenerates data from the Radio Unit (RU), as well as from the Distributed Unit (DU), making it indispensable for environments where O-RAN standards are not fully implemented, and communication is needed to the CU or the RIC. Similarly, it translates RAN control messages to the RU/DU/CU entities that are not fully O-RAN compliant, such as new technologies like ISAC or Cell-Free Massive MIMO.

Co-funded by the European Union
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.
6GSNS

There are currently two primary trends in O-RAN deployment: systems that are fully compliant with O-RAN standards, featuring fully functional O-RAN interfaces, and those that are not fully compliant, lacking complete O-RAN interface integration or providing extensions that are not specified by O-RAN. For the latter, the telemetry framework is crucial as it facilitates the necessary translation between non-compliant interfaces and the rest of the O-RAN ecosystem.

Additionally, raw data that come from the radio environment or abstracted data that come from the real-world through sensing capabilities, can be leveraged on the TGW and fed to the Near-RT RIC and Non-RT RIC for decision making and RAN control. As an example, the TGW can collect input values from the Kafka bus and publish the calculated output metrics with the same timestamp, so other xApps use them independently of its source.

Moreover, the telemetry framework simplifies the development process by abstracting the complexities of O-RAN interfaces for xApp/rApps application developers. Instead of requiring developers to interact directly with the intricate details of O-RAN interfaces, and the specific type of messages from RUs and DUs, the framework offers processed and relevant information directly to the RIC. This abstraction not only eases the development process but also ensures that developers can focus on creating applications without delving into the underlying O-RAN infrastructure details. This also applies for the controlling messages, as the TGW can handle abstracted messages such as simplified handover messages written in some xApp, while TGW will generate the appropriate message structure needed for lower layer components.

The technical requirements for the telemetry framework within O-RAN architecture focus on ensuring comprehensive data management and control across the network's various components, specifically categorized into two main areas: telemetry messages or metric messages, and control messages.

- Telemetry Messages or Metric Messages. This category encompasses the capabilities required to handle and process data from the lower RU and CU components of the O-RAN standardization framework. Essentially, it involves the collection, preprocessing, and abstraction of metrics reported by the User Equipment (UE) to the CU, DU, and RU. These metrics must be processed within the telemetry framework in a manner that is not only efficient but also aligned with both O-RAN and 3GPP specifications. The preprocessing of these metrics is critical for transforming raw data into usable metrics on the xApp/rApps ensuring that the data is standardized and interoperable across the various elements of the O-RAN ecosystem.
- Control Messages. The second key area, control messages, deals with integrating the operational capabilities of the radio components within the network. This includes functions such as activating or deactivating cells, providing inter-frequency handovers, modifying physical layer characteristics, and managing mobility aspects. These control operations are vital for the dynamic management of the network, allowing for real-time adjustments to optimize performance and efficiency based on current network conditions and demands. The execution of these control messages is contingent upon the specific capabilities and configurations of the radio equipment in use, necessitating a flexible and adaptable telemetry framework that can accommodate a wide range of operational scenarios.

Together, these two areas of technical requirements underline the complexity and sophistication needed in the telemetry and control framework to ensure seamless data integration and network control within the O-RAN architecture.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

3.2. Heterogeneous data sources and harmonisation

Harmonisation of data is a critical requirement when building a digital twin that considers several data sources and types. To ensure the inclusion of such data within the digital twin, it is essential to incorporate an interoperability layer. This layer, comprising software components that facilitate communication and data exchange between diverse systems, applications, and components, acts as a bridge between different technologies, protocols, and formats. Hence, this interoperability layer is a key element in the solution design, to enable seamless integration and interaction.

The interoperability layer must consist of open APIs (Application Programming Interfaces) and protocols that provide a standardised interface for communicating with other systems. Interoperability will allow developers and other actors (apps, other data sources, etc.) to integrate different components and services with the digital twin, even if they are built with various technologies or use different data formats.

In terms of interoperability, several initiatives should be considered, such as the 3GPP, ETSI (specifically, the ENI and NFV working groups), ITU standards, NETCONF YANG Model (based on the RFC 6020) and oneM2M, which provides a comprehensive architecture that includes data harmonisation parts.

Nonetheless, the main focus of this chapter is to enable data harmonisation, so we will focus on modelling the information and making sure it is available in an interoperable way. With this in mind, the interoperability layer ensures that data is consistent and accurate, even from different sources, by providing translation and mapping services to normalised ones (SmartDataModels). Overall, it is a crucial component, as it allows for a standard interface for communication and data exchange and helps to ensure that data is accurate, consistent, and secure.

This framework for interoperability can support the construction of a robust system capable of integrating with other systems quickly and rapidly, causing as little disruption as possible to existing systems. These integrations happen between existing systems, requiring them to have some technology capable of communicating with the digital twin.

In this way, the different systems will be integrated into the solution in a secure and controlled manner while the integrations to be carried out will also serve as the main channel for receiving data from external applications, providing them to the information systems or external applications where they are needed.

From all of the available Data Harmonisation methods, which include ETL (Extract, Transform, and Load), Data Cleaning, Master Data Management, Data Fusion, and Schema Mapping, our solution focuses on ETL. This is because ETL is the most flexible of them all. By using tools like Apache Airflow, we can script any of the other methods into the ETL pipeline when needed.

In addition to what has already been mentioned, the solution's interoperability mechanisms allow the system to add new data sources regardless of whether they are formatted for the desired protocols or in other formats. In this way, creating a robust, repeatable, predictable and, above all, scalable ecosystem is possible. For integration purposes, two different approaches can be taken:

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

- ***Job-Based (poll-based)*** - Where jobs (code scripts) are developed for integration, mediated with a scheduler that determines the periodicity of the integration. These jobs can be changed programmatically to accommodate new types of data or new conversions and can be reused for different integrations, using good DRY (Don't Repeat Yourself) programming practice. Tools like Apache Airflow can be used for this purpose. In a 6G context, the poll-based approach can be used in two main ways: (1) Recurrent Polling: A cron-like script would be set up to poll network devices or compute units at regular intervals (e.g., every 5 minutes) to collect data; and (2) Database Polling: By polling a distributed database storing 6G network logs to extract and transform the data for analytics and network management;
- ***Push Based*** - In this case, specific integrators will be made available to expose the protocols needed to integrate the messages from the devices. The device must register on the platform and then initiate communication via, for example, MQTT, gRPC, HTTP or Websockets. Two possible real scenarios in the 6G context would be: (1) Telemetry Data from IoT Devices, where 6G-enabled IoT devices would push telemetry data to a central data processing system (the ETL pipeline) via MQTT (Message Queuing Telemetry Transport) or HTTP POST requests; and (2) Real-Time Monitoring of network health: where 6G network components push performance metrics and logs to a Kafka topic. The ETL pipeline would be subscribed to such a topic for the below-mentioned process.

With this in mind, we aim for an ETL-based process to implement the necessary changes to the heterogeneous data sources, making them all one common standard. It should be noted that the data is only harmonised after following through this ETL pipeline.

The high-level ETL process will include the common iterations, more specifically:

- **Extract** where the data reaches the ETL pipeline. Here, several connectors will be implemented for the different technologies needed in the scope of the NDT application. For example, gRPC, HTTP, MQTT, and Kafka connectors will be considered;
- **Transform** where the data is cleaned, standardised and converted into the standard format, for example, the Smart Data Model. Here, the necessary mapping and transformation-based techniques (based on scripting) will be applied since we assume that the input formats are heterogeneous and might not follow the output specifications and data modelling the project will push forward. In this sense, this step is the most important one and represents the core of the ETL overall solution to create an interoperability point for the NDT;
- **Load** is where the data is stored on the target system, in this case, a data space. This data space is the only place where the data is harmonised in the common data model and standard that the project has regarded and chosen. Therefore, the architecture for the data flow needs to be put in this central place: the NDT data space.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

3.3. Data collection and distributed storage

In the world of network management and operations, the challenge of gathering and organising network data for faults, performance, and events is an ever-present challenge. This data has always been of vital importance in ensuring the optimal functioning of networks, given its direct correlation with network performance, network configuration for optimisation and usage trends for capacity planning.

Furthermore, the dynamic nature of the telecommunications industry perpetually ushers in newer interfaces and protocols that are commonly juxtaposed with existing legacy systems. This combination of old and new poses a significant barrier to the integration and harmonisation of performance management data. In addition, newly defined interfaces for AI/ML data collection for RIC such as O1, E2 and A1 offer an array of options, implemented to varying degrees, all using different transport protocols (TLS, TCP, SCTP).

In addition to newer standards, networks have been deployed with performance monitoring agents where devices incorporate built-in performance monitoring agents that collect and report performance data directly from the device hardware or software. Flow-based telemetry protocols like gRPC (gRPC Remote Procedure Calls) and gNMI (gRPC Network Management Interface) provide high-resolution, real-time telemetry data from network devices. In comparison, Simple Network Management Protocol (SNMP) has been long used to monitor the operational status, performance metrics, and configuration parameters of base stations deployed in GSM, UMTS, LTE, and 5G networks. This includes monitoring radio resource utilisation, throughput, signal quality, and hardware health. Remote Network Monitoring (RMON) is an extension of SNMP that enables remote monitoring and management of network devices. RMON provides additional capabilities beyond standard SNMP, including packet capture, traffic analysis, and event notifications whereas many devices offer APIs for programmatic access to performance data.

The challenge to collect multiple data sources is more pertinent today due to the promise of AI/ML to consume this data to automatically control and optimise the most valuable of resources, the RAN.

More recently, the E2 interface was defined to collect data from specific O-RAN Elements (DU, CU-UP and CU-CP). E2 Nodes provide services with access to messages and measurements and to enable control of the E2 Node from the Near-RT RIC. RIC Services include Report, Insert, Control, Policy, and Query. For each RIC Service, different types of data can be grouped as a style. A given E2 node may support many styles for each RIC service. These styles represent predefined groupings or formats of data tailored to specific operational needs, ensuring efficient and flexible management by the RIC. Each E2 node advertises their RAN functions which may support different parts / capabilities of the E2SM (E2 Service Model). RAN Function#1 may only support event trigger styles 1 and 2 while RAN Function#2 may support event trigger styles 3 and 4. Each parameter is identified through <parameter ID, parameter value> pairs where parameter ID to corresponding ASN.1 mapping is specified in existing 3GPP specifications and results in interoperability issues mapping between vendors. Furthermore, E2 defines overlapping Service Modules that are not uniformly adopted by the industry.

Even though E2 measurements are identified based on 3GPP defined measurement names [20], with reporting granularity is from 1ms, the E2 standards and adoption are maturing slowly.

Within the O-RAN Architecture, the Service Management and Orchestration (SMO) Framework is arguably gaining most momentum, as it aligns nicely with current network and

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

management infrastructure. It communicates over the non-realtime O1 Interface, which includes implementation of Fault, Configuration, Accounting, Performance, Security (FCAPS) functions, File management and Software management functions of both physical and virtual network element functions. The non-realtime Performance Management (PM) component aligns with existing 3GPP defined network counters and as such are readily available from existing element management systems. Hence, large parts of the O1 PM data sets can be collected from both ORAN and legacy systems, over the same non-realtime frequency.

When the non-RT RIC determines that measurement data is needed it interacts with the SMO's OAM Functions to collect measurement data from the network. The SMO generates a PM Job and performs the PM Job control operations accordingly, and the network elements decide if the job is acceptable or not, in other words, it is ultimately the network element who decides whether the measurement data collection task can be established or not. All O-RAN PM job control by the SMO is transacted via the O1 Interface.

However, to enrich the Machine Learning (ML) model and to obtain more powerful predictions, the NDT may require external data sources (such as news, events, sensors, and weather), all of which requires mediation of both legacy and O-RAN interfaces into relevant in-memory databases and/or Kafka to create a Data Space Based around datacenter tools such as KubeFlow.

In 6G-TWIN, a blueprint for data collection of heterogeneous data types is a crucial component for reinforced network learning and access to the rich data sources are required to be available in a consistent manner [25].

The overall data collection process must account for various types of storage and data handling methodologies, as outlined in Figure 6. Specifically, this includes "Data Indexing", "Real-time Data Storage", and "Historical Data Storage". "Data Indexing" refers to organizing and tagging data to enable efficient retrieval and query processing. "Real-time Data Storage" involves capturing and storing data as it is generated, allowing immediate access for real-time analysis and decision-making. "Historical Data Storage" is the archiving of data for long-term storage and analysis, providing insights into trends and patterns over time.

Moreover, the data storage requirements shall be addressed in a distributed manner. In this sense, each of the network components can have its own data storage point which would ultimately act as the data fabric that feeds the overarching data space as defined by Section 2. Moreover, such a solution shall be able to provide an interconnected fabric with a central storage piece which acts as the main (local) storage point for this instance of an NDT.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

4. Governance, Privacy, and Security Requirements

The integration of network digital twins and 6G network brings out new threat landscapes. This threat landscape is revolving around the data space governance, and the associated security and privacy requirements. A large amount of data from various sources of physical network infrastructure is collected, processed, and analysed by these network digital twins to mirror it. This situation involves implementing strict security measures to maintain data integrity, security and privacy in highly dynamic 6G environment.

4.1. Governance

4.1.1. General principles

Governance of NDT data space is required to ensure an efficient and fair data sharing. It requires to ensure the preservation of rules (ethical, regulatory, policy, code of conduct), business models in a fair and secured manner for the users (trustworthy transactions). It is also supervised by a data space governance authority that ensure the compliance of the aforementioned points.

To ensure fairness, transparency, and compliance throughout the process, a variety of documents are necessary to establish the governance of the data space and its respective use cases:

1. **Data Space governance framework:** This encompasses a set of principles, standards, policies, agreements, and practices governing the management and operation of a data space, including both business and technological aspects. It also outlines procedures for enforcement and dispute resolution. The Data Space Governance Authority defines this framework.
2. **Contractual framework (constitutive agreement):** This primary legally binding document establishes the foundation for a Data Space Use Case or the entire Data Space. The Data Space Governance Authority at the data space level and the Data Space Use Case Orchestrators at the use case level define this framework.
3. **Accession agreement:** Governs the admission of parties into either the Data Space or a specific Data Space Use Case. The Data Space Governance Authority at the data space level and the Data Space Use Case Orchestrators at the use case level define this agreement.
4. **Code of conduct:** Consists of commonly accepted norms facilitating cooperation among parties within a Data Space or Data Space Use Case. The Data Space Governance Authority at the data space level and the Data Space Use Case Orchestrators at the use case level define this code.
5. **Access and usage policies:** Outlines the terms under which Data & Service Providers grant rights to use their products to Service Providers and/or End Users. The involved parties define these policies together.

This process involves multiple stakeholders and layers, addressing various topics across the governance levels:

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

- **Standards:** Determining technical, legal, and business standards.
- **Code of conduct:** Establishing the applicable code.
- **Roles and responsibilities:** Assigning roles and associated responsibilities.
- **Use cases:** Identifying permissible use cases.
- **Infrastructure:** Selecting tools and processes to meet governance, legal, and business requirements.
- **Business and pricing models:** Defining allowable business models. Proposing data valuation beyond monetisation: business, societal, environmental, etc.
- **Accession rules:** Specifying conditions for joining the data space**.**
- **Data and service usage policies:** Establishing conditions for using available data sets and services.
- **Conflict resolution policies.**

It is necessary to have a **Data space authority** that will ensure compliance with the code of conduct, regulation compliance (e.g., privacy) and fairness (respecting a business model). In the context of national NDT, the national telecommunication regulator can play this role.

4.1.2. 6G-TWIN principles

The Data Management Plan (DMP) made in the context of 6G-TWIN already provides insights regarding legal, ethics and usage that could be ported to the NDT data space we are defining.

All the important topics could not be covered at the time of writing this deliverable (i.e., Month 6 of the project), notably because of the wide range of potential data providers and content. Without a complete overview we cannot go beyond the DMP recommendation. It will be detailed in Work Package 4 (WP4) of the project, when the case study will bring the concrete case, where we can define more properly business models (data value chain, contract), accession rules and roles and responsibilities of the different stakeholders.

4.2. Security

The simulation framework is coupled to several components: the closed-loop framework for the management, the model repository, and one or several simulators running basic and functional models. All these software components may be executed on different servers, maybe servers of different datacenters hosted by different companies. Therefore, secure communication between these components must be realized.

Digital Twin (DT) emerged in the beginning of 2000's [26] and they are developing in several fields including network which brings out Network Digital Twin (NDT) [27]. There are efforts to standardize the architecture, requirements of DTs to assist interoperability. The Internet Engineering Task Force (IETF) [28], The International Telecommunication Union (ITU) [8], Industrial Internet Consortium (IIC) [29] focuses on building the concepts and reference architecture for NDTs while IEEE, ISO and IEC have continuing standardization studies. Also, working groups in digital twin consortium [30] targeting the security and reliability of DTs. There are related studies to investigate the risk and threats, identify attack surfaces and analysing the requirements for a secure DT [31]. Apart from the existing studies, a threat analysis and

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

security requirements investigation considering the 6G specific use cases performed in this study.

Threat analysis is a way to find the vulnerabilities with their possible effects to make an overall system reliable and secure. There are standardized frameworks to conduct threat analysis in each system and architecture. In here, STRIDE threat modelling (Spoofing, Tampering, Repudiation, Information disclosure, Denial of service, Elevation of privilege) [32] is applied to the proposed high level 6G-Twin approach that can be seen in Figure 7 below.

To be able to apply STRIDE threat modelling (i) the assets in the overall system (ii) the type and the capabilities of the adversaries (iii) threat surface and the analysis on these assets are identified.

As it is shown in the overall architecture in Figure 7 below, the system has physical and digital assets. These assets are used to collect data from the network, evaluate the network's state and control related components of the network. In this process, there are two-way data flows between each of the following pairs: (i) physical network to NDT (ii) physical network to network's application layer (iii) NDT to network's application layer and (iv) internal elements through NDT.

For these assets and data flows adversaries can be categorized considering their access to the NDT. Thus, there might be insider (I), outsider (O) and targeted/hacker (H) attackers that might want to compromise the NDT. These attackers have different level of skills and knowledge about the digital twin. A threat analysis considering the attackers' roles will be given after describing the threat surfaces for the proposed network digital twin approach.

The threat surface includes all the weaknesses, routes, and techniques that an adversary can use to get into networks, sensitive data, or carry out cyberattacks. Threat surfaces of the NDT are listed below considering the system architecture and description as well as the data flow through this architecture. Simultaneously, a threat analysis for each threat surface is performed below considering the assets, adversaries, and threat surfaces (list of potential attackers / STRIDE threat categories).

Data storage and repository in NDT	- An attacker might have an unauthorized access to the data related with the network or network functionality that are twinned to NDT. (I, O, H / T) - The data might be exfiltrated by the targeted attacker. (I, O, H / I) - Critical and confidential data, functional models, ML models from the NDT might be copied, stolen, and compromised (I, H / I).
Any sensor or software from data collection process	- An adversary might spoof the data traffic and flow between the NDT and physical network. (H, O / S)
Physical network components	- An adversary could tamper with the physical assets (H, O / T)

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

Data flow through between the physical network assets and DT (data gathering, control/actuation and synchronization processes)	- An attacker might sniff and compromise the synchronization between NDT and physical network and compromise the decisions, actuations and management of the network (H / D, I) - The data flow might be tampered with or sniffed by an attacker (H, O / T, I) - A flooding attack might be performed in the data gathering channel to create a denial-of-service (H, O / D)
Network management models (ML/AI models or simulations) and functions	- An attacker might try to gain access to the network management infrastructure and functions. The attacker might actively compromise the management of the network (might compromise the configurations, topology data, security/access policies etc.) Also, the attacker might steal/exposure private network management process and data. (H, I / S, T, R, I, D, E)
Application layer interface (APIs)	- An adversary might tamper with the application layer data via interfaces and network access (H, O / T) - Request flows and the data through application layer interfaces might be sniffed by an attacker. During this sniffing confidential information such as physical processes/network functions might be gathered by an attacker (H, O / I) - Attacker might cause a denial-of-service attack on the NDT and its components (H, O / D)

4.3. Privacy

NDT has a lot of potential for managing and orchestrating networks; however, its security and privacy issues might significantly impede its widespread use [33]. NDT may lead to a number of security flaws and privacy violations due to the:

- protection of network's vital infrastructures
- extensive personal and network related data gathering
- extensive sharing of data among different digital twins

The data and data flow mentioned above must traverse numerous networks, software programs, and applications throughout its full lifecycle in order to provide services. The data collected from the various parts of the network might have a previously unseen degree of detail and high synchronization frequency to preserve a digital replica of physical network. Also, it is susceptible to inherit all existed vulnerabilities such as phishing, eavesdropping, poisoning etc.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

since NDT is based on a number of cutting-edge technologies for its service offerings [34]. Shortly, the emerging NDT environment might give rise to entirely unanticipated risks including virtual-reality compounded threats which has a risk potential for novel attacks and privacy-related concerns.

In this section, the privacy landscape and threats for proposed 6G-TWIN architecture below (Figure 7), and related countermeasures are discussed. These privacy concerns are expected to mostly rise up from: (i) data collection and storage processes, (ii) data flow through the physical or digital entities, (iii) access mechanisms and authentication related issues [35]. The use cases that might create privacy concerns are grouped and listed below.

Continuous acquisition of network data: A great deal of network data is expected to be gathered in NDT with an unprecedented degree of detail. This private and sensitive data could be misused or leaked.

Continuous flow of network data: The data might traverse between physical network and digital twin as well as between different digital twins. Also, data might traverse inside the NDT itself. These intra/inter-NDT/physical-virtual twin interactions create new attack surfaces.

- An insider attacker could make use of her/his privilege within the NDT and its resources to obtain sensitive data. This sensitive data could be used to illegally manage the NDT and physical network.
- An outsider attacker could eavesdrop/gain access to sensitive data and could manipulate the NDT services, network orchestration.

Accountability of the NDT services and data usage: It is challenging to identify the compromised entities in the NDT and promptly execute accountability owing to the intricacy of NDT services. This situation might encourage a malicious party to violate privacy concerns.

Data misuse: Authorized service providers might inadvertently reveal privacy-sensitive data. Also, an unauthorized attacker might sell the accessed data for financial advantages.

Collaborative learning and model aggregation*:* There is an information leakage risk due to the cloud/edge servers that might store the training data due to the nature of collaborative learning paradigm.

Adherence to regulations in NDTs*:* The authorized service providers must comply with privacy and confidentiality regulations such as GDPR and consider these regulations when storing/collecting and processing the network's data. Therefore, consent of network users and service consumers might be required to main regulations related to privacy.

Opacity of service providing*:* Conflicts and opacity over resources and services on the service provider's side might create a concern on the part of the service consumer that privacy is being violated. It may also lead service providers to be more flexible about their privacy prioritization.

There are various mechanisms that are suggested [36] to increase the privacy in NDT data space. These countermeasure mechanisms can be declaimed under three headings [37] (i) blockchain based solutions (ii) federated learning for collaborative learning scenarios (iii) privacy preserving computation methods.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

Blockchain based solutions	- Blockchain integrated federated learning architecture [38] - Blockchain integrated model update scenario to keep local and global model updates synchronized [39] - Blockchain integrated data verification and sharing processes.
Federated learning for collaborative learning scenarios	- The privacy of edge devices and different parts of a network can be safeguarded in NDTs by not sharing any critical data - Federated learning architecture could be used to conduct training process locally at edge devices especially in wireless networks that includes personal privacy-critical data at edge devices [40]
Privacy preserving computation methods	- Secure multi-party computing - Differential privacy - Homomorphic encryption

However, the blockchain based solutions are still dubious due to the shortcomings of blockchain itself such as computation, scalability, and storage.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

5. Connecting AI-Native NDT to the Overall Infrastructure

In this section, we aim to define the process of integrating AI-native NDTs with the broader infrastructure of a potential 6G network. The focus will be on the communication aspects, which are fundamental in ensuring seamless interaction between the NDT and the physical network.

This section builds on the foundational concepts discussed in previous sections, providing a comprehensive understanding of how AI-native NDTs interact with the physical network and other simulation frameworks.

The high-level architecture presented in Figure 7 depict three main layers (green, blue and orange). The green part of Figure 7 represents the physical network and the operations that are performed on it. This includes user equipment (UE), the surrounding environment that can influence the propagation patterns, the RAN and the core network. The blue part comprises the NDT models and related data. The NDT model area will consist of a set of models to best and dynamically represent the network elements (basic models) and their behaviour (i.e., functional models). These models will rely on the harmonised data collected, and will be developed to optimise network planning, management and control issues, as well as to provide meaningful KPIs and metrics to operators. These DT models will be federated in several layers, allowing complex, interconnected and dynamic topologies to be imagined depending on the applications to be served. Finally, the orange part consists in an open source, secured and federated simulator.

We aim to highlight the methods used to ensure data integrity and synchronization between the different architectural layers as well as between several NDT instances. The subsequent subsections will detail the communication to and from the physical network, interactions with simulation frameworks, and the requirements for connecting multiple DTs.

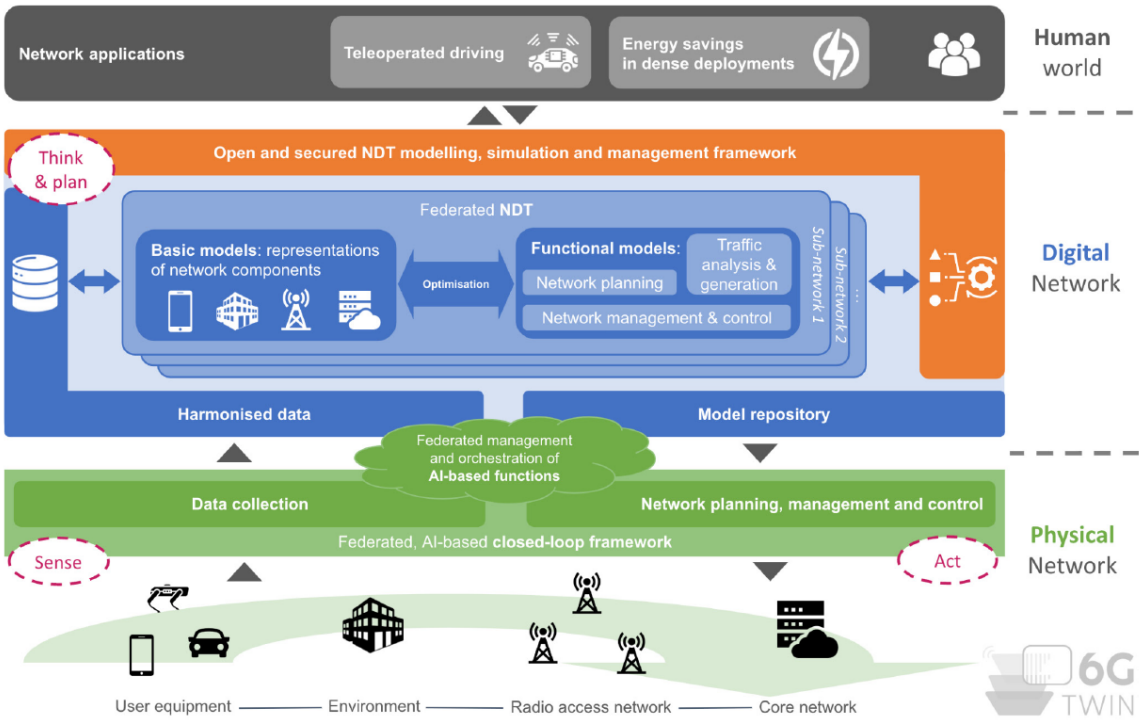

Figure 7. High-level overview of 6G-TWIN NDT architecture

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

5.1. Communication from the physical network: real and synthetic interfaces

Effective network monitoring and telemetry of network elements (see section 3) to obtain information related to KPIs, such as bandwidth usage, signal strength, latency, and error rates, are critical for ensuring modern telecommunication systems' robust performance and reliability and for the NDT to maintain an accurate and up-to-date model of the physical network. These processes involve collecting, processing, and analysing data from various segments of the network, including the UE, RAN, edge devices, transport network, core network, and cloud infrastructure. Real interfaces are the actual physical connections that transmit data across these segments, providing real-time insights into network performance, traffic patterns, and potential issues. On the other hand, synthetic interfaces are virtual representations or simulations that mimic real network behaviours, enabling enhanced analysis and predictive capabilities inside an NDT without impacting the live network inside

Complementing the monitoring/sensing elements, edge devices located at the network's periphery gather and preprocess data before sending it to the NDT. This reduces latency and bandwidth usage by processing data closer to where it is generated. Cloud infrastructure provides scalability and computational power for extensive data analysis and storage. Big Data platforms like Hadoop and real-time analytics engines such as Apache Flink and Storm are utilized to store and analyse large volumes of network data, enabling the NDT to provide immediate insights and predictions at large scale.

In the scope of 6G-Twin we will need to rely on synthetic data that are very close to reality. It should be used in the case of early design of our NDT (i.e., before having the data) or when only having partial data (e.g., due to privacy, regulation compliance or trade-secret). The Viavi RAN Scenario Generator (RSG) provides a rich set of RAN data about the performance of the RAN under given load scenarios. The datasets provide a snapshot of RAN performance, per UE, per Cell, per Network Slice at any given point in time. The data is made available to management systems using standards-based and file-based interfaces, that are consistent with the communication to and from the physical network to management systems. These interfaces include the O1 and E2 explained in Section 2.2.2.

Today, the physical attributes of the RAN can be imported into RSG (Cell Site GPS Coordinates, Height, Tilt, Azimuth, Antenna Configuration, Radiation Patterns) in addition to the actual topology (e.g. via GIS Maps) enabling the static component of NDT to be modelled. However, this data can be complemented by real network data measurements – actual UE mobility patterns, actual performance metrics and actual traffic patterns. As mentioned in section 3.1, these data sets are accessed by means of various interfaces, probes, and APIs. The RSG tool will also act as a data provider in the Network Data space conforming to the proposed Smart Data Model. Conversion of network data via a wide range of interfaces (EMS, SNMP, E2, O1 etc) into InfluxDB and/or Kafka based data space, enabling homogeneous access to heterogeneous data sources.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

5.2. Communication to the physical network

The NDT concept involves creating a virtual replica of a physical network to simulate, analyse, and optimize its performance. The last aspect, optimization, requires communication from the NDT towards the network elements to interact with them. This interaction relies on robust interfaces and architectural components facilitating seamless data exchange, but there are three main ways that 6G-TWIN propose to achieve it: a) **direct interaction**, where the digital component directly controls the physical network elements [41] (e.g., a controller in the NDT can directly interface with the physical network infrastructure), b) **semi-direct interaction**, where there is an interface between the digital and physical twin, but the digital counterpart has the logic functionality, and the physical one translates it into changes in the infrastructure (e.g., a virtual SDN controller sends the actions to be performed, and the physical one simply translates it into actual network commands) [42] [43], and c) **indirect interaction** via a logic functionality created using the NDT but deployed on the physical twin [44] [45]. (e.g., a NDT to be used as sandbox for cybersecurity or training AI/ML models for controlling 5G deployments)

Before we describe them in more detail, it is important that while in an NDT there may be many network elements that can be present in the DT, from decision-makers to sensors and actuators, we will focus mainly on the first ones, which provide the capabilities to perform network management, e.g., orchestrators and controllers. Orchestrators manage the lifecycle of network services, ensuring efficient allocation of network resources to meet service requirements. Examples include the NFV Orchestrator (NFVO) and Cloud Orchestrator. Controllers manage specific network domains by enforcing policies, routing traffic, and ensuring optimal performance. The RAN Intelligent Controller (RIC) manages radio access networks, while the SDN Controller oversees software-defined networks.

Direct interactions: In this approach, decision-making network elements inside the NDT (DT elements) send optimized configuration and control commands to the physical twin elements it controls in the DT environment based on their analysis and simulation outcomes. Closed-loop automation is a key mechanism here, where the NDT continuously monitors the network, simulates various scenarios, and dynamically adjusts network configurations. Event-driven actions also play a role, with the NDT triggering specific actions in response to events such as network congestion or security breaches directly on the elements that need to be changed.

Semi-direct interaction: In this approach, decision-making network elements inside the NDT send optimized configuration and control commands back to their physical twin elements based on their analysis and simulation outcomes, but it is the physical element that performs changes (e.g., controlling) in the network. Like the direct interaction approach, closed-loop automation is a key mechanism here, where the NDT continuously monitors the network and simulates various scenarios, but instead of controlling directly the network elements to be changed, it provides feedback to physical twin orchestrators and controllers that will take care of dynamically adjust network configurations. In other words, the physical twin will only translate the insights and predictions generated by the NDT and execute control commands sent by it, enabling dynamic adjustments to network configurations and performance parameters, such as adjusting signal power, rerouting traffic, or activating/deactivating network components to optimize performance and efficiency.

Indirect interaction: The NDT provides a controlled environment to create, train in case of AI/ML algorithms, and validate the algorithms with the logic to orchestrate and control the network. By simulating various network scenarios and conditions, the NDT can help develop robust algorithms capable of making intelligent decisions in real-time. For example, an AI-based auto scaler can be trained to predict network traffic patterns and automatically scale

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

resources up or down to meet demand, ensuring optimal performance and resource utilization. Once trained, these algorithms can be integrated directly with network controllers and orchestrators. The AI-based autoscaler, for instance, can be embedded within the NFV Orchestrator to manage virtual network functions dynamically. Similarly, ML models can be deployed within the SDN Controller to optimize routing decisions based on real-time network conditions.

Although these three approaches are possible, the complexity of the NDT varies depending on each. As in the simpler description NDT is a software component, we can compare each approach to develop the NDT based on quality software attributes [46] as described in the following table.

Table 5. Quality software attributes of an NDT depending on how it interacts with the physical twin.

Interaction Type	Reliability	Usability	Efficiency	Scalability	Flexibility	Maintainability
Direct	High - Continuous monitoring and real-time adjustments ensure consistent network performance.	Moderate - Direct control requires specialized knowledge and understanding of the physical network elements and include it in the NDT system.	High - Immediate adjustments optimize resource use and network performance.	Moderate - Scaling may be limited by direct control mechanisms and integration complexities in the NDT.	Moderate - Limited flexibility due to tight integration with physical elements.	Low - High coupling between NDT and physical elements makes maintenance complex and error-prone.
Semi-direct	High - Feedback loops enhance reliability by involving physical orchestrators and controllers.	High - Easier to use as physical elements handle control, allowing operators to focus on the logic behind the controller/orchestrator behind the NDT.	Moderate - Indirect control may introduce some latency but still efficient due to continuous monitoring.	High - Easier to scale as physical elements can independently handle changes based on NDT feedback.	High - Greater flexibility as physical elements interpret and implement NDT commands.	Moderate - Easier to maintain than direct interaction due to clear separation of control and execution layers.
Indirect	Moderate - Reliability depends on the accuracy and robustness of developed algorithms.	High - User-friendly interfaces for training algorithms using the NDT and deploying them simplify usage.	High - Algorithms can optimize resource utilization dynamically based on real-time data.	High - Highly scalable as algorithms can be easily replicated and distributed across the network.	High - Highly flexible as AI/ML models can be retrained and updated to adapt to changing conditions.	High - Easier to maintain as changes can be made to algorithms independently of the physical network elements.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

5.3. Communication with the simulation framework in 6G-TWIN

The simulations should allow that the real-world physical network, the NDT, and the NDT simulation interact. The data basis of these interactions is the data space. The "simulation framework" itself has no direct connection to the data space or the physical network that is modelled by the NDT. However, the physical network can interact with simulations: Real-world data is used by the closed-loop optimization framework to generate realistic configurations that are given to the simulation framework to simulate them, illustrated in Figure 8.

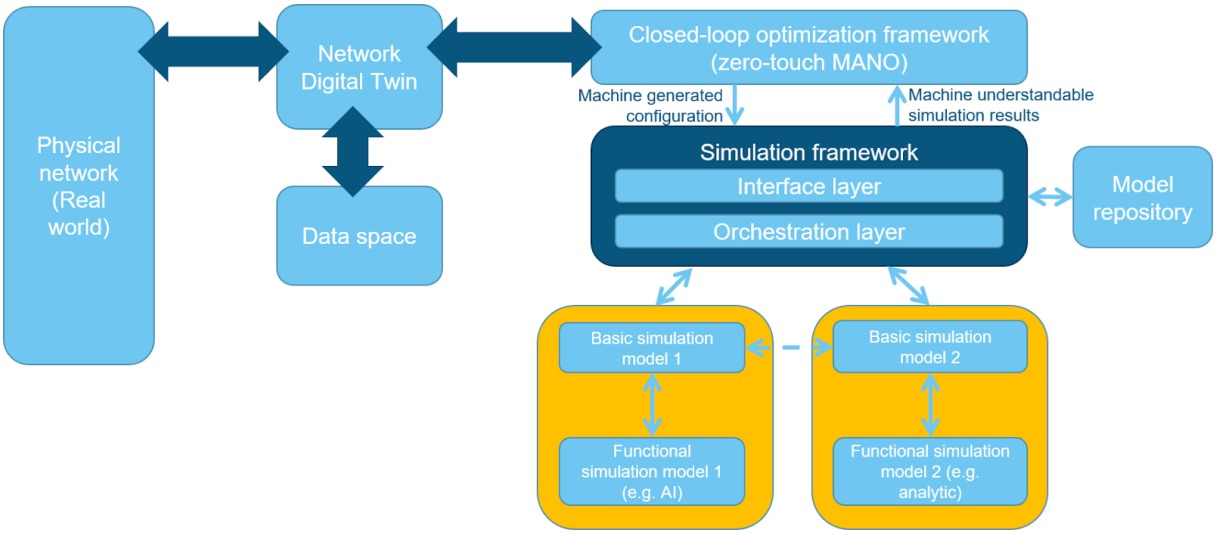

Figure 8. Interaction of the closed-loop framework, simulation framework and the simulators with the "real world"

Besides the interaction with the physical network, APIs are required for the interaction of the simulation framework and the closed-loop optimization framework, the model repository, and the simulators. The requirements about these interfaces are described in deliverable D1.1, section 3.45.

The simulation framework does not directly access the data space, because the closed-loop optimization framework selects the information that is necessary to define a simulation that fulfils a concrete goal in the context of the NDT. The simulation framework receives this simulation definition from the closed-loop optimization framework and manages the simulators that are required to execute this simulation. Simulation results are returned to the closed-loop optimization framework that draws conclusions from these results, e.g. in an optimization loop that contains a lot of simulations. If these conclusions result in relevant data for the NDT itself, this information is stored by the closed-loop optimization framework in the data space of the NDT.

5.4. NDT Creation and Deployment Modes

Digital twins are virtual representations of physical entities, processes, or systems synchronized with real-world counterparts. These digital replicas enable organizations to

Co-funded by the European Union
Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.
6GSNS

monitor, analyse, and optimize the performance of their physical assets, ultimately improving efficiency, reducing costs, and enhancing decision-making. However, their complexity will not only depend on what they are representing but also how they are built (e.g., single vs. a composition of multiple NDTs) and its deployment (e.g., centralized vs. federated) [47].

5.4.1. Single Instance Deployment

As seen in Section 3, a single digital twin can provide valuable insights into the behaviour and performance of a specific network. It simplifies the way a holistic view on the network is built. It does not necessarily work on a centralised data storage: in the context of the NDT data space some data may remain the property of their respective data owner and consumed on demand by the NDT, but from the execution point of view their components are orchestrated and deployed in a centralized way [48].

A single instance NDT works for local NDT (i. e., working on a small area), or with a very specific concern (e.g., an NDT for RAN, or an NDT for the edge computing infrastructure). Nevertheless, for large case it could not scale computationally [49], due to GDPR compliance, respecting data sovereignty of each party, or simply computationally is not able to run in a central computing unit [50].

5.4.2. Multiple Instance Deployment

The true power of digital twins emerges when they are integrated into a federated network of digital twins. This federated network of digital twins allows for the seamless exchange of data and the coordination of actions across multiple systems, enabling a more comprehensive understanding of the overall ecosystem [47], [51]. Indeed, it could relate to other digital twins pertaining to domain influencing indirectly the Network such as city building, topography, etc. In the context of 6G networks, the concept of a federated network digital twin becomes particularly relevant since they are expected to be highly complex, with a diverse array of interconnected devices, infrastructure, and services. A federated network digital twin for 6G would create a virtual representation of this intricate system, allowing network operators to simulate, test, and optimize the network's overall performance. By integrating data from various sources, such as sensors, network elements, and applications, the federated network digital twin can provide a holistic view of the network's behaviour, enabling real-time monitoring, predictive maintenance, and proactive optimization. This, in turn, can lead to improved network reliability, reduced downtime, and enhanced user experiences, all of which are critical for the successful deployment and operation of 6G networks.

Within all the different ways to implement the federated digital twin, one element should remain the same: relying on a data space to exchange data amongst the NDTs. Indeed, the communication between NDT should follow the rules defined for data exchanges (see Data Space, Section 2) notably in terms of data security, privacy, sovereignty.

We can envision multiple approaches for federation: orchestration or choreography [52].

- **Orchestration** is having federation with one orchestrating NDT that calls from view coming from sub-NDT; This federation can also be hierarchical orchestration [53].
- **Choreography** consists in relying on a distributed view of the Network amongst NDT. They could exchange information, through data connectors when required, notably

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

when a modification on one network part could impact the other or when it needs external information involved in another NDT.

In Figure 9, we show an example of orchestration with two sub-digital twins acting in a federation of NDT, relying on data connector to ensure that the orchestrating NDT is compliant with the mentioned rules.

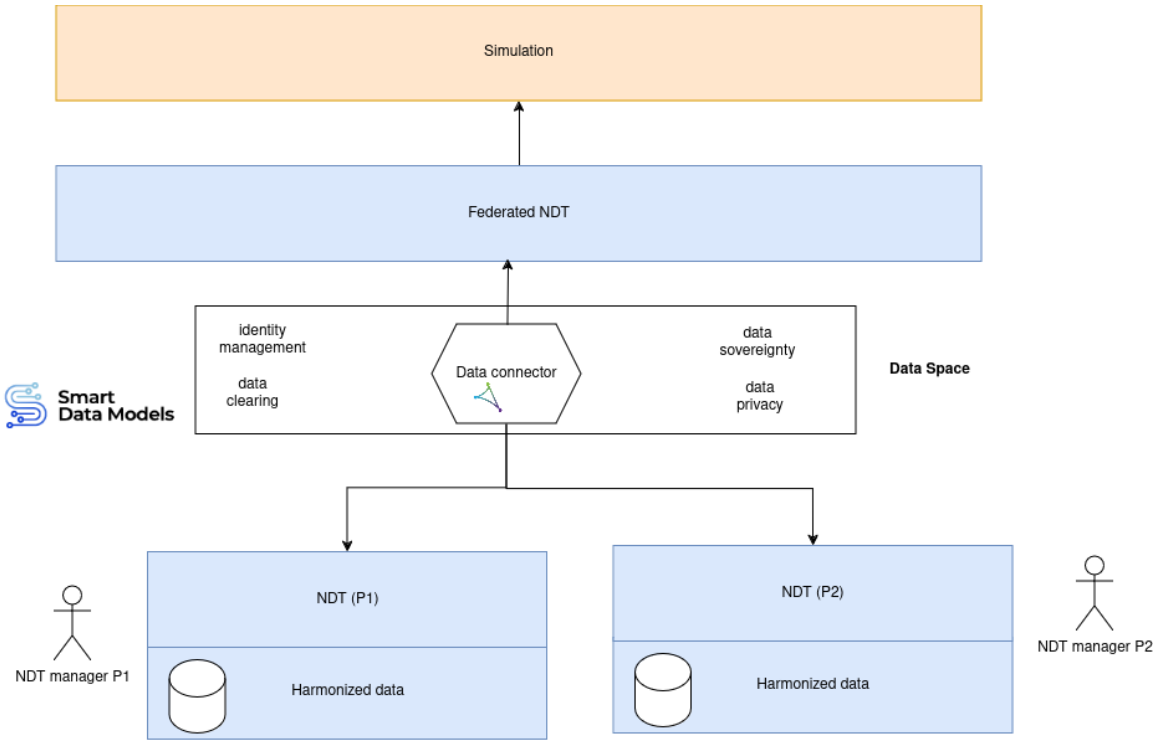

Figure 9. Multi-DT federation (orchestration) with Data space.

5.4.3. The 6G-TWIN case

As seen in Section 3, in the context of 6G-TWIN a reference data model and a process for interoperability will be a made in the context of this WP2. It will make the federation of NDT and their interconnection a lot easier to deal with. Such an interface is possible thanks to the context management API and data information models. Several APIs within a bus layer shall follow and comply with reference implementations of such concepts, such as, Open and Agile Smart City consortium's (OASC) Minimal Interoperability Mechanisms (MIM), a set of practical capabilities based on open technical specifications that allow entities to replicate and scale solutions worldwide. From OASC specifically MIM1: Context Information and MIM2: Data Models to be regarded as the models to follow.

Such solutions provide the technical basis for acquiring and implementing data platforms and complete solutions worldwide. With the work on Smart Data Models for Network and keeping Data Spaces at the centre of data sharing, we will ensure that the necessary interfaces and standard information blocks are available and can communicate.

Therefore, to solve the federated-DT aspect, each DT shall adhere to common standards regarding interfaces, interoperability, and semantic representation of the information. This

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

does not mean that all underlying systems must be adapted. Instead, a standard federation layer, which can be operated by an NDT orchestrator, should appear and be available. With these solutions in place, we can bootstrap a standard DT for all our data points.

All in all, the upstream flow of data shall include the following aspects, which derive from Figure 5 in Section 2.4 in more detail:

- Data points. These are already in place in a given environment (API, web services, RIC, other apps or sources). These data points are better explained in the document (namely in the Technical Requirements section), but they can include (1) distributed databases, (2) gRPC connections, (3) HTTP connections, (4) Kafka brokers, and (5) MQTT messages.
- NDT interoperability framework. This will harmonise and standardise the local information on the local database. It represents the main ETL (see Section 3) pipeline of the system, which details and implements all the necessary translation and transformation tooling to enable the standard data model to arise.
- NDT standard API. This will create the common entry point for the multi-DT aspect of the work to appear, making it a seamless integration.
- NDT Data Space. The Data Space of the NDT which acts as a distributed data point in the multi-DT ecosystem. It should contain the harmonised information of the devices, services, and any other computing processes.
- Identity & Access Management. This component will hold the policies for access to the data and enable federated access from the Federated DT to the different NDTs and their data.
- Federated DT API. This API allows to communicate all the NDT information and provide the necessary information for the Simulation works that will be conducted.

When preparing the WP3 contributions focused on simulation, we will initially opt for a simple deployment mode involving a single NDT. The primary goal will be to design efficient simulation solutions and develop communication APIs (see Section 5.3) that can support more complex deployments in the future, such as utilizing multiple simulations within a federation of NDTs. The availability of synthetic data (e.g., RSG generator presented in Section 5.1) will simplify data ownership issues, eliminating the need for a complex data-space ecosystem and allowing local NDTs to process their own data independently.

During the implementation of the case studies in WP4, we will need to select one of the various deployment variants of our NDT. Based on data availability and case complexity, we will likely start with a single NDT approach. This initial focus on simplicity will help manage the complexity of implementing essential elements such as basic models, functional models, simulators, and network controllers. However, this approach will not hinder the potential expansion of the cases. If we need to handle different networks (e.g., cross-border situations), each with its own NDT, we will adopt a federation approach, whether through orchestration or choreography. In such scenarios, data space governance will be extended to facilitate data exchange between NDTs, ensuring seamless integration and communication across the network.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

6. Conclusions

This deliverable provides a comprehensive framework for data governance, privacy, and harmonization within 6G-TWIN. Data is the fuel of any Digital Twin. NDT does not step away from this, but it does have its own specificities in terms of data sharing (e.g., many providers, confidentiality of data) and potential size (e.g., can be worldwide). This document sets the groundwork for effective and secure data management in future 6G networks by addressing key aspects such as data collection, telemetry, harmonization, privacy, and security. The main contributions of this document are summarised as follows.

Section 2 introduces the concept of Data Spaces in the context of 6G NDTs. It establishes a common standard Smart Data Model (currently in incubation [24]) based on existing 3GPP and O-RAN standards for 5G networks, to ensure interoperability and consistency across the 6G NDT. This model will be further exploited in the upcoming activities of the project, with the objective to serve as a reference for the other existing or future projects dealing with NDTs.

Section 3 Details the technical requirements for data collection, distributed storage, and telemetry framework for 6G NDT. The challenges of handling heterogeneous data sources and the need for harmonisation are also addressed.

Section 4 addresses data governance principles, that are essential for managing data within the 6G-TWIN framework, with a strong focus on security and privacy measures to ensure compliance with regulatory standards.

Section 5 defines the components and processes necessary for creating a coherent and sustainable NDT data space, and its integration with the overall NDT architecture, which is crucial for the long-term viability and success of NDTs in 6G networks. It specifies the integration and communication with the physical network and with simulation tools. Furthermore, it defines alternative deployment models: from a single NDT to a federation of NDTs.

The previous key aspects define the implementation ecosystem of a Network Data Space. The Network Data Space is the central component that will allow to fuel the NDT with data respecting all the constraints, rules, regulation necessary to ensure the sustainability of the NDT in a longer term, i.e., ensuring exchange fairness for data providers as well as data sources up and running. It will help to organise all the data exchanges between different parties including the need to manage a federation of NDTs.

This deliverable is only the first step towards establishing data management in 6G-TWIN project. The incubated Smart Data Model initialises the work to be done in the context of Task 2.2 (Basic Model). The complete data harmonization process could then start when have the full cartography of data handled by the Basic model. Further developments will also be performed through the respective progress of WP1, WP2 and WP3.

Synthetic data will be largely used to validate/evaluate research proposing from WP2 (functional models), WP3 and WP4. Those synthetic data will be easier to manage from data ownership point of view because member of the consortium produce them. But this will be used to prepare the data governance. As synthetic will emulate the real data, we can annotate synthetic data with potential data sources, data providers and thus prepare under which rules (regulation, value for exchange) a potential data space could work. Preparing a proper data space will also allow to integrate synthetic data coming from other sources.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

A specific data governance will be established in the context of the WP4 when effectively dealing with data in each case study. The Data connector, central to Data space will be specified and eventually an existing implementation (e.g., the Eclipse Data Connector) selected when knowing more precisely the data providers. Similarly, more details about data collection and telemetry will be given through implementation of the case study.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

References

[1] 'EU Data Strategy 2020 – Data Economy'. [Online]. Available: http://dataeconomy.eu/eu-data-strategy-2020/

[2] M. T. Delgado, 'Eclipse Dataspace Components | projects.eclipse.org'. [Online]. Available: https://projects.eclipse.org/projects/technology.edc

[3] 'International Data Spaces (IDS)', International Data Spaces. [Online]. Available: https://internationaldataspaces.org/

[4] 'Smart Data Models'. [Online]. Available: https://smartdatamodels.org/

[5] 'Dataspace Protocol 2024-1 | Dataspace Protocol | IDS Knowledge Base'. [Online]. Available: https://docs.internationaldataspaces.org/ids-knowledgebase/v/dataspace-protocol/

[6] 'eclipse-tractusx/identity-trust'. Eclipse Tractus-X, Mar. 20, 2024. Accessed: Jun. 11, 2024. [Online]. Available: https://github.com/eclipse-tractusx/identity-trust

[7] T. Usländer et al., 'Symbiotic Evolution of Digital Twin Systems and Dataspaces', Automation, vol. 3, no. 3, pp. 378–399, Aug. 2022, doi: 10.3390/automation3030020.

[8] 'ITU-T REC-Y.3090 (2022/02), Digital twin network – Requirements and architecture'.

[9] '3GPP TS 28.622 version 16.4.0 Release 16; Generic Network Resource Model (NRM); Integration Reference Point (IRP); Information Service (IS)'. Accessed: May 23, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/128600_128699/128622/16.04.00_60/ts_128622v160400p.pdf

[10] '3GPP TS 32.401 version 17.0.0 Release 17; Telecommunication management; Performance Management (PM); Concept and requirements'. Accessed: May 27, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/132400_132499/132401/17.00.00_60/ts_132401v170000p.pdf

[11] '3GPP TS 23.501 version 16.6.0 Release 16; 5G; System architecture for the 5G System (5GS)'. Accessed: Apr. 11, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/123500_123599/123501/16.06.00_60/ts_123501v160600p.pdf

[12] 'O-RAN ALLIANCE e.V'. [Online]. Available: https://www.o-ran.org/

[13] 'TS 38.300 - version 16.2.0 Release 16: NR; NR and NG-RAN Overall Description'.

[14] '3GPP TS 38.470; 5G; NG-RAN; F1 general aspects and principles, v15.5.0'. Accessed: Jun. 17, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138400_138499/138470/15.05.00_60/ts_138470v150500p.pdf

[15] '3GPP TS 38.460: NG-RAN; E1 general aspects and principles'. Accessed: Jun. 17, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138400_138499/138460/16.01.00_60/ts_138460v160100p.pdf

[16] 'ORAN-WG3.E2GAP.0-v02.02, O-RAN Working Group 3; Near-Real-time RAN Intelligent Controller Architecture & E2 General Aspects and Principles v02.02'.

[17] 'O-RAN-WG10.O1-Interface-v07.00: O-RAN Operations and Maintenance Interface Specification v07.00'.

[18] 'O-RAN-WG6.CAD-v03.00: Cloud Architecture and Deployment Scenarios for O-RAN Virtualized RAN v03.00'.

[19] 'O-RAN-WG2.A1GAP-v02.03: O-RAN Working Group 2; A1 interface: General Aspects and Principles v02.03'.

[20] 'O-RAN-WG4.MP.0-v09.00: O-RAN Alliance Working Group 4; Management Plane Specification v09.00'.

[21] '3GPP TS 24.501 version 16.5.1 Release 16; Non-Access-Stratum (NAS) protocol for 5G System (5GS); Stage 3'. Accessed: Jun. 27, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/124500_124599/124501/16.05.01_60/ts_124501v160501p.pdf

[22] '3GPP TS 38.413 version 16.7.0 Release 16; 5G; NG-RAN; NG Application Protocol (NGAP)'. Accessed: Jun. 27, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138400_138499/138413/16.07.00_60/ts_138413v160700p.pdf

[23] '3GPP TS 38.331 version 16.1.0 Release 16; 5G; NR; Radio Resource Control (RRC); Protocol specification'. Accessed: Jun. 11, 2024. [Online]. Available: https://www.etsi.org/deliver/etsi_ts/138300_138399/138331/16.01.00_60/ts_138331v160100p.pdf

[24] 'incubated/CROSSSECTOR/6G-TWIN at master · smart-data-models/incubated', GitHub. [Online]. Available: https://github.com/smart-data-models/incubated/tree/master/CROSSSECTOR/6G-TWIN

[25] Y. Lu et al., 'Machine Learning for Synthetic Data Generation: A Review'. arXiv, May 02, 2024. Accessed: Jun. 28, 2024. [Online]. Available: http://arxiv.org/abs/2302.04062

[26] D. M. Botín-Sanabria, A.-S. Mihaita, R. E. Peimbert-García, M. A. Ramírez-Moreno, R. A. Ramírez-Mendoza, and J. de J. Lozoya-Santos, 'Digital Twin Technology Challenges and Applications: A Comprehensive Review', Remote Sens., vol. 14, no. 6, Art. no. 6, Jan. 2022, doi: 10.3390/rs14061335.

[27] P. Almasan et al., 'Digital Twin Network: Opportunities and Challenges'. arXiv, Jan. 07, 2022. Accessed: Apr. 11, 2024. [Online]. Available: http://arxiv.org/abs/2201.01144

[28] C. Zhou et al., 'Digital Twin Network: Concepts and Reference Architecture', Internet Engineering Task Force, Internet Draft draft-irtf-nmrg-network-digital-twin-arch-04, Oct. 2023. Accessed: Jun. 11, 2024. [Online]. Available: https://datatracker.ietf.org/doc/draft-irtf-nmrg-network-digital-twin-arch-04

[29] 'Digital Twins for Industrial Applications', Industry IoT Consortium. [Online]. Available: https://www.iiconsortium.org/digital-twins-for-industrial-applications/

[30] 'Digital Twin Consortium'. [Online]. Available: https://www.digitaltwinconsortium.org/

[31] A. Raghuramu et al., 'Network Digital Twins: A Threat Analysis', in 2023 IEEE International Conference on Communications Workshops (ICC Workshops), May 2023, pp. 733–739. doi: 10.1109/ICCWorkshops57953.2023.10283573.

[32] R. Khan, K. McLaughlin, D. Laverty, and S. Sezer, 'STRIDE-based threat modeling for cyber-physical systems', in 2017 IEEE PES Innovative Smart Grid Technologies Conference Europe (ISGT-Europe), Sep. 2017, pp. 1–6. doi: 10.1109/ISGTEurope.2017.8260283.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.

6GSNS

[33] Y. Wu, K. Zhang, and Y. Zhang, 'Digital Twin Networks: A Survey', IEEE Internet Things J., vol. 8, no. 18, pp. 13789–13804, Sep. 2021, doi: 10.1109/JIOT.2021.3079510.

[34] M. Sanz Rodrigo, D. Rivera, J. I. Moreno, M. Àlvarez-Campana, and D. R. López, 'Digital Twins for 5G Networks: A Modeling and Deployment Methodology', IEEE Access, vol. 11, pp. 38112–38126, 2023, doi: 10.1109/ACCESS.2023.3267548.

[35] 'ELEGANT: Security of Critical Infrastructures With Digital Twins | IEEE Journals & Magazine | IEEE Xplore'. [Online]. Available: https://ieeexplore.ieee.org/document/9499077

[36] D. Chen, D. Wang, Y. Zhi, and Z. Han, 'Digital Twin for Federated Analytics Using a Bayesian Approach | IEEE Journals & Magazine | IEEE Xplore'. [Online]. Available: https://ieeexplore.ieee.org/document/9491072

[37] Y. Wang, Z. Su, S. Guo, M. Dai, T. H. Luan, and Y. Liu, 'A Survey on Digital Twins: Architecture, Enabling Technologies, Security and Privacy, and Future Prospects', IEEE Internet Things J., vol. 10, no. 17, pp. 14965–14987, Sep. 2023, doi: 10.1109/JIOT.2023.3263909.

[38] L. Jiang, H. Zheng, H. Tian, S. Xie, and Y. Zhang, 'Cooperative Federated Learning and Model Update Verification in Blockchain-Empowered Digital Twin Edge Networks', IEEE Internet Things J., vol. 9, no. 13, pp. 11154–11167, Jul. 2022, doi: 10.1109/JIOT.2021.3126207.

[39] S. Son, D. Kwon, J. Lee, S. Yu, N.-S. Jho, and Y. Park, 'On the Design of a Privacy-Preserving Communication Scheme for Cloud-Based Digital Twin Environments Using Blockchain', IEEE Access, vol. 10, pp. 75365–75375, 2022, doi: 10.1109/ACCESS.2022.3191414.

[40] W. Sun, N. Xu, L. Wang, H. Zhang, and Y. Zhang, 'Dynamic Digital Twin and Federated Learning With Incentives for Air-Ground Networks', IEEE Trans. Netw. Sci. Eng., vol. 9, no. 1, pp. 321–333, Jan. 2022, doi: 10.1109/TNSE.2020.3048137.

[41] G.-P. Liu, 'Control Strategies for Digital Twin Systems', IEEECAA J. Autom. Sin., vol. 11, no. 1, pp. 170–180, Jan. 2024, doi: 10.1109/JAS.2023.123834.

[42] D. R. R. Raj, T. A. Shaik, A. Hirwe, P. Tammana, and K. Kataoka, 'Building a Digital Twin Network of SDN Using Knowledge Graphs', IEEE Access, vol. 11, pp. 63092–63106, 2023, doi: 10.1109/ACCESS.2023.3288813.

[43] M. Kherbache, M. Maimour, and E. Rondeau, 'When Digital Twin Meets Network Softwarization in the Industrial IoT: Real-Time Requirements Case Study', Sensors, vol. 21, no. 24, Art. no. 24, Jan. 2021, doi: 10.3390/s21248194.

[44] S. Vakaruk, A. Mozo, A. Pastor, and D. R. López, 'A Digital Twin Network for Security Training in 5G Industrial Environments', in 2021 IEEE 1st International Conference on Digital Twins and Parallel Intelligence (DTPI), Jul. 2021, pp. 395–398. doi: 10.1109/DTPI52967.2021.9540146.

[45] A. Mozo, A. Karamchandani, M. Sanz, J. I. Moreno, and A. Pastor, 'B5GEMINI: Digital Twin Network for 5G and Beyond', in NOMS 2022-2022 IEEE/IFIP Network Operations and Management Symposium, Apr. 2022, pp. 1–6. doi: 10.1109/NOMS54207.2022.9789810.

[46] 14:00-17:00, 'ISO/IEC 25002:2024, Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — Quality model overview and usage', ISO. [Online]. Available: https://www.iso.org/standard/78175.html

[47] C. R. Vergara, G. Theodoropoulos, R. Bahsoon, W. Yanez, and N. Tziritas, 'Federated Digital Twins as an Enabling Technology for Collaborative Decision-Making', in Proceedings of the 38th ACM SIGSIM Conference on Principles of Advanced Discrete Simulation, Atlanta GA USA: ACM, Jun. 2024, pp. 67–68. doi: 10.1145/3615979.3662152.

[48] Md. S. Khan et al., 'Centralized Digital Twin driven Supply Chain for Sustainable Smart Manufacturing Plants', in 2022 International Conference on Computational Modelling, Simulation and Optimization (ICCMSO), Dec. 2022, pp. 370–373. doi: 10.1109/ICCMSO58359.2022.00077.

[49] L. Adreani, P. Bellini, M. Fanfani, P. Nesi, and G. Pantaleo, 'Smart City Digital Twin Framework for Real-Time Multi-Data Integration and Wide Public Distribution', vol. 12, 2024.

[50] J. Jeon, B. Jeong, and Y.-S. Jeong, 'Intelligent Resource Scaling for Container-Based Digital Twin Simulation of Consumer Electronics', IEEE Trans. Consum. Electron., vol. 70, no. 1, pp. 3131–3140, Feb. 2024, doi: 10.1109/TCE.2023.3320174.

[51] M.-S. Baek, 'Digital Twin Federation and Data Validation Method', in 2022 27th Asia Pacific Conference on Communications (APCC), Oct. 2022, pp. 445–446. doi: 10.1109/APCC55198.2022.9943622.

[52] M. Autili, A. Di Salle, F. Gallo, C. Pompilio, and M. Tivoli, 'CHOReVOLUTION: Service choreography in practice', Sci. Comput. Program., vol. 197, p. 102498, Oct. 2020, doi: 10.1016/j.scico.2020.102498.

[53] H. K. Ravuri, M. T. Vega, J. van der Hooft, T. Wauters, and F. De Turck, 'A Scalable Hierarchically Distributed Architecture for Next-Generation Applications', J. Netw. Syst. Manag., vol. 30, no. 1, p. 1, Sep. 2021, doi: 10.1007/s10922-021-09618-4.

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or Smart Networks and Services Joint Undertaking. Neither the European Union nor the granting authority can be held responsible for them.